



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification: C12N 15/52, C12N 5/10, C12N 9/00, C12N 15/63, C12P 17/06	A2	(11) International Publication Number: WO 00/22139 (43) International Filing Date: 20 April 2000 (20.04.2000)
--	-----------	---

(21) International Application Number: PCT/US99/23535 (22) International Filing Date: 11 October 1999 (11.10.1999) (30) Priority Data: 198 46 493.2 09 October 1998 (09.10.1998) DE (60) Parent Application or Grant GESELLSCHAFT FÜR BIOTECHNOLOGISCHE FORSCHUNG MBH (GBF) [/]; (). BRISTOL-MYERS SQUIBB COMPANY [/]; (). BEYER, Stefan [/]; (). BLOECKER, Helmut [/]; (). BRANDT, Petra [/]; (). CINO, Paul, M. [/]; (). DOUGHERTY, Brian, A. [/]; (). GOLDBERG, Steven, L. [/]; (). HOFLE, Gerhard [/]; (). MUELLER, Rolf- Joachim [/]; (). REICHENBACH, Hans [/]; (). BEYER, Stefan [/]; (). BLOECKER, Helmut [/]; (). BRANDT, Petra [/]; (). CINO, Paul, M. [/]; (). DOUGHERTY, Brian, A. [/]; (). GOLDBERG, Steven, L. [/]; (). HOFLE, Gerhard [/]; (). MUELLER, Rolf-Joachim [/]; (). REICHENBACH, Hans [/]; (). SANTUCCI, Ronald, R. : ().	Published
--	------------------

(54) Title: DNA SEQUENCES FOR ENZYMATIC SYNTHESIS OF POLYKETIDE OR HETEROPOLYKETIDE COMPOUNDS
 (54) Titre: SEQUENCES D'ADN DESTINEES A LA SYNTHÈSE ENZYMATIQUE DE COMPOSÉS À BASE DE POLYKETIDES
 OU D'HETEROPOLYKETIDES

(57) Abstract

The invention consists of: (1) cloned *Sorangium cellulosum* polyketide synthase (PKS) biosynthetic cluster DNA; and (2) the nucleotide sequence and predicted protein coding sequences of the cloned DNA. The invention can be used for, but not limited to: (a) increasing yields of PKS product in *Sorangium cellulosum* (e.g., by amplification or genetic modification of the epothilone gene cluster or its component parts); (b) increasing yields of polyketide product in a heterologous system by transfer of the epothilone gene cluster or its component parts, which may be followed by amplification or genetic modification of the PKS gene cluster or its component parts; (c) modification of the polyketide product chemical structure in either *Sorangium cellulosum* or a heterologous host (e.g., by genetic modification of the epothilone gene cluster or its component parts; and (d) for the detection of genes and gene products involved in making polyketides or related molecules in other organisms (e.g., by hybridization or complementation assays). DNA sequence and analysis is presented for the following cosmids and plasmids: A2 cosmid; the pEPOcos6 region (overlapping of pEPOcos6 and pEPOcos7); pEPOcos8 cosmid; A5 cosmid; Sau4 (10 kb plasmid).

(57) Abrégé

L'invention concerne: (1) un ADN biosynthétique cloné en grappe de polykétide synthase (PKS) de *Sorangium cellulosum*; et (2) la séquence nucléotidique et les séquences protéiques codantes prévues de l'ADN cloné. L'invention peut avoir les applications suivantes (sans caractère limitatif): (a) augmentation de la production de PKS chez *Sorangium cellulosum* (p.ex., par l'amplification ou la modification génétique de la grappe de gènes épothilone ou de ces parties constitutives); (b) augmentation de la production du produit polykétide dans un système hétérologue par le transfert de la grappe de gènes épothilone ou de ces parties constitutives, qui peut être suivie par l'amplification ou la modification génétique de la grappe de gènes PKS ou des ses parties constitutives; (c) modification de la structure chimique du produit polykétide soit chez *Sorangium cellulosum* soit chez un hôte hétérologue (p.ex., par l'amplification ou la modification génétique de la grappe de gènes épothilone ou de ces parties constitutives); et (d) détection de gènes et de produits géniques participant à la fabrication de polykétides ou de molécules correspondantes dans d'autres organismes (p.ex., par des dosages à hybridation ou à complémentation). La séquence d'ADN et l'analyse sont présentées pour les cosmides et les plasmides suivants: cosmide A2; région pEPOcos6 (se chevauchant avec pEPOcos6 et pEPOcos7); cosmide pEPOcos8; cosmide A5; Sau4 (plasmide 10 kb).

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/52, 9/00, 15/63, 5/10, C12P 17/06	A2	(11) International Publication Number: WO 00/22139
		(43) International Publication Date: 20 April 2000 (20.04.00)

(21) International Application Number: PCT/US99/23535

(22) International Filing Date: 11 October 1999 (11.10.99)

(30) Priority Data:
198 46 493.2 9 October 1998 (09.10.98) DE

(71) Applicants (for all designated States except US):
GESELLSCHAFT FÜR BIOTECHNOLOGISCHE
FORSCHUNG MBH (GBF) [DE/DE]; Mascheroder Weg 1,
D-38124 Braunschweig (DE). BRISTOL-MYERS
SQUIBB COMPANY [US/US]; P.O. Box 4000, Princeton,
NJ 08543-4000 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): BEYER, Stefan [DE/DE];
Mascheroder Weg 1, D-38124 Braunschweig (DE).
BLOECKER, Helmut [DE/DE]; Mascheroder Weg 1,
D-38124 Braunschweig (DE). BRANDT, Petra [DE/DE];
Mascheroder Weg 1, D-38124 Braunschweig (DE). CINO,
Paul, M. [US/US]; 4 Crest Drive, Bound Brook, NJ 08805
(US). DOUGHERTY, Brian, A. [US/US]; 10 Rosemary
Lane, Killingworth, CT 06419 (US). GOLDBERG, Steven,
L. [US/US]; 25 Compton Court, Basking Ridge, NJ 07920
(US). HOFLE, Gerhard [DE/DE]; Mascheroder Weg 1,
D-38124 Braunschweig (DE). MUELLER, Rolf-Joachim

[DE/DE]; Mascheroder Weg 1, D-38124 Braunschweig
(DE). REICHENBACH, Hans [DE/DE]; Mascheroder Weg
1, D-38124 Braunschweig (DE).

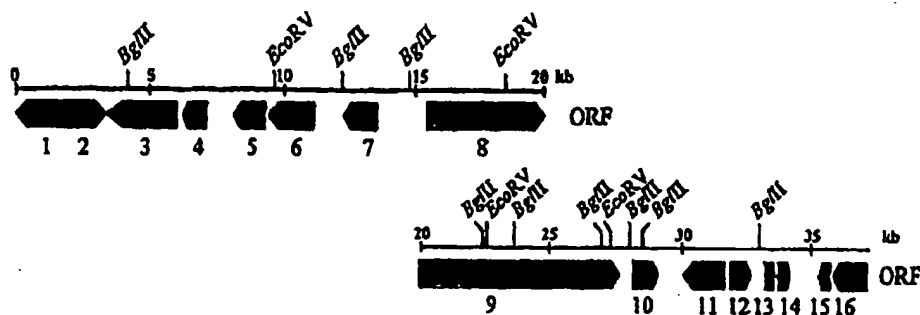
(74) Agents: SANTUCCI, Ronald, R. et al.; Kane, Dalsimer,
Sullivan and Levy, LLP 20th floor, 711 Third Avenue, New
York, NY 10017 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG,
BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE,
ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP,
KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA,
MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU,
SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG,
US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE,
LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM,
AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT,
BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU,
MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM,
GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished
upon receipt of that report.

(54) Title: DNA SEQUENCES FOR ENZYMATIC SYNTHESIS OF POLYKETIDE OR HETEROPOLYKETIDE COMPOUNDS



(57) Abstract

The invention consists of: (1) cloned *Sorangium cellulosum* polyketide synthase (PKS) biosynthetic cluster DNA; and (2) the nucleotide sequence and predicted protein coding sequences of the cloned DNA. The invention can be used for, but not limited to: (a) increasing yields of PKS product in *Sorangium cellulosum* (e.g., by amplification or genetic modification of the epothilone gene cluster or its component parts); (b) increasing yields of polyketide product in a heterologous system by transfer of the epothilone gene cluster or its component parts, which may be followed by amplification or genetic modification of the PKS gene cluster or its component parts; (c) modification of the polyketide product chemical structure in either *Sorangium cellulosum* or a heterologous host (e.g., by genetic modification of the epothilone gene cluster or its component parts; and (d) for the detection of genes and gene products involved in making polyketides or related molecules in other organisms (e.g., by hybridization or complementation assays). DNA sequence and analysis is presented for the following cosmids and plasmids: A2 cosmid; the pEPOcos6 region (overlapping of pEPOcos6 and pEPOcos7); pEPOcos8 cosmid; A5 cosmid; Sau4 (10 kb plasmid).

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LJ	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

D scription

5

10

15

20

25

30

35

40

45

50

55

5

10

15

**DNA sequences for enzymatic synthesis of polyketide or
heteropolyketide compounds**

10

20

The present invention relates to DNA sequences for enzymatic synthesis of polyketide or heteropolyketide compounds produced by the bacterium *Sorangium cellulosum*.

25

15 Background and introduction

30

This patent application describes DNA sequences for the enzymatic synthesis of polyketide and/or heteropolyketide structures synthesized by the myxobacterium *Sorangium cellulosum*. Several of these compounds have known cytotoxic, immunosuppressive, antibiotic and fungicidal biological activity, with the epothilones having been most studied and characterized. The fermentation of large quantities of secondary metabolites from microorganisms, especially from myxobacteria, is a time consuming and difficult process that often involves complications (e.g. contamination, low product yield, difficult isolation and purification). Therefore it would be advantageous to use a well-characterized organism for such fermentations. After cloning of the desired biosynthetic genes one could create such an organism via genetic engineering and manipulate the biosynthesis of the compound. Identified sequences

40

45

30

50

55

5

2

can be cloned into optimized expression vectors and generate recombinant cell lines that overproduce polyketide structures.

10

Polyketide synthases (PKS) and non-ribosomal peptide synthetases (NRPS) represent macromolecular and multifunctional enzymes which are characterized by a modular architecture. PKS condenses activated carbonic acids (usually acetate and propionate) and reduce the resulting 2-keto acid intermediates stepwise in a fatty acid biosynthesis-like fashion. Responsible for each reaction step is a specific domain that recognizes, activates, condenses and reduces the carbonic acid. Depending on the presence of these domains in the corresponding modules, every reduction stage can occur in the final product (Rawlings, Nat. Prod. Reports 14, 523-556 [1997]; for a review, see Chem. Rev. 27, 2463-2760 [1997]). A typical example for the biosynthesis of a polyketide is the macrolide antibiotic erythromycin (Staunton and Wilkinson, Chem. Rev. 27, 2611-2630 [1997]). NRPSs are also modular enzymes and condense via peptide bonds amino acids to low molecular weight bioactive substances like bacitracin or tyrocidin. Typical domains of these systems activate the amino acid and condense it with the growing peptide chain. Methylations, epimerisations and modifications via additional protein domains are possible (Stachelhaus and Marahiel, FEMS Microbiol Lett. 125, 3-14 [1995]). Both types of enzymes (NRPS and PKS) share the modular organization of the proteins in which specific catalytic domains are responsible for recognition, activation, condensation and modification of the single elongation units. The growing chain of amino acids and/or carbonic acids is extended through the action of one module adding one unit. The domains of each module carry the active centers responsible for the enzymatic steps of the biosynthesis.

5

15

20

25

15

30

35

40

45

50

55

5

3

10

15

20

25

30

35

40

45

50

55

Little is known about the biosynthesis of biologically active polyketides and polypeptides from myxobacteria. Fragments of the biosynthetic gene clusters of soraphen and saframycin have been described from *Sorangium cellulosum* So ce26 and *Myxococcus xanthus*, respectively (Schupp et al., J. Bacteriol. 177, 3673-3679 [1995] and Pospiech et al., Microbiology 141, 1793-1803 [1995]). We have constructed genomic libraries of the epothilone producer *Sorangium cellulosum* So ce90. Gene probes based on PKS and PS genes were used to isolate recombinant cosmids, which were then sequenced and characterized. Several unique pathways containing PKS, PS, or a combination of both types of genes were identified, demonstrating that this organism is potentially a rich source of novel bioactive compounds.

A subject of the present invention is therefore to provide DNA sequences according to claim 1 the expression products of which perform or are involved in the enzymatic biosynthesis, mutasynthesis or partial synthesis of polyketide or heteropolyketide compounds. The DNA sequences may be inserted into well known and optimized expression vectors by common techniques of molecular biology, thus allowing transformation, selection and cloning of cells, which cells are then capable of synthesizing polyketide or heteropolyketide compounds by fermentation. Using an overproducing clone allows the desired polyketide or heteropolyketide compounds be easily produced and recovered in high amounts. Further, knowledge of the localization of regulatory DNA segments and individual structural genes allows "site-directed mutagenesis" using common techniques for genetic engineering, and thus construction of optimized enzymes ("protein engineering") for fermentative synthesis of polyketide or heteropolyketide compounds.

5

4

10

The invention thus further relates to a recombinant expression vector according to claim 16, cells transformed there-
with according to claim 17 and to a process for enzymatic biosynthesis, mutasynthesis or partial synthesis of polyketide or
5 heteropolyketide compounds according to claim 23.

15

Preferred and/or advantageous embodiments of the present invention are subject-matter of the subclaims.

20

In brief, the invention consists of (1) cloned *Sorangium*
cellulosum polyketide synthase (PKS) and/or peptide synthetase
10 (PS) biosynthetic cluster DNA and (2) the nucleotide sequence and predicted protein coding sequences of the cloned DNA. The invention can be used for, but not limited to, (a) increasing
25 yields of PKS product in *Sorangium cellulosum* (e.g., by amplification or genetic modification of the epothilone gene cluster
15 or its component parts), (b) increasing yields of polyketide and/or peptide synthetase product in a heterologous system by
30 transfer of the corresponding gene cluster or its component parts, which may be followed by amplification or genetic modification of the PKS and/or PS gene cluster or its component
20 parts, (c) modification of the polyketide and/or peptide synthetase product chemical structure in either *Sorangium cellulosum* or a heterologous host (e.g., by genetic modification of
35 the corresponding gene cluster or its component parts) and (d) for the detection of genes and gene products involved in making
40 polyketides or related molecules in other organisms (e.g., by hybridization or complementation assays). DNA sequence and
25 analysis is presented for the following cosmids and plasmids:

45

- A2 cosmid as defined in claim 6
- the pEPOcos6 region (overlapping of pEPOcos6 and pEPOcos7)
30 as defined in claim 7

50

55

5

5

10

- pEPOcos8 cosmid as defined in claim 10
- A5 cosmid as defined in claim 12
- Sau4 (10 kb plasmid) as defined in claim 14

15

5 The invention is now described in more detail by examples and
for illustration only. The examples are not to be construed as
any limitation of the scope.

20

10

Figure 1 is a restriction map of one of the DNA sequences of
the present invention (cosmid A2 insert) indicating also the
localization of regulatory DNA segments and the individual
structural genes ("open reading frames" or ORFs) 1 to 16.

25

15

Figure 2 shows the open reading frames found on pEPOcos6 region

30

DNA sequence data from A2 cosmid are as defined in claim 6.

35

Table 1 correlates ORFs 1 to 16 found on A2 cosmid with the re-
spective biological function (Regulators, Enzymes).

40

45

50

55

5

6

Table 1

10

15

20

25

30

35

40

45

	gene/function	position
ORF 1	regulatory element	1666 - 1
ORF 2	regulatory element	1605 - 3338
ORF 3	acyl-t-RNA synthetase	6100 - 3398
ORF 4	monooxygenase	7110 - 6374
ORF 5	amino transferase	9590 - 8433
ORF 6	L-dopa decarboxylase	11393 - 9855
ORF 7	oxidoreductase	13656 - 12712
ORF 8	polyketide synthase	15374 - 18984
ORF 9	polypeptide synthetase	20003 - 27889
ORF 10	peptidase	28251 - 29402
ORF 11	regulatory element	31720 - 30401
ORF 12	sigma factor	31982 - 32932
ORF 13	regulatory element	33128 - 33613
ORF 14	regulatory element	33661 - 34007
ORF 15	transcription regulator	35611 - 35255
ORF 16	signal transduction	37856 - 35730

50

55

5

7

Working Examples

10

A. Construction of a *Sorangium cellulosum* cosmid library5 1. Isolation of genomic DNA from *S. cellulosum* So ce90

15

a. *Sorangium cellulosum* So ce90 was spread onto solid CA-2 agar and incubated at 30°C for 5-7 days. CA-2 agar is prepared by autoclaving 18 g Bacto-agar (Difco Laboratories, Detroit, MI) in 800 ml dH₂O for 20 min at 121°C and cooling to 50-55°C in a water bath. The following filter-sterilized solutions are added to the agar: 20% (w/v) glucose, 50 ml; Solution A (7.5% [w/v] KNO₃, 7.5% K₂HPO₄), 10 ml; Solution B (1.5% [w/v] MgSO₄·7H₂O), 10 ml; Solution C (0.2% [w/v] CaCl₂·2H₂O, 0.15% [w/v] FeCl₃), 10 ml; 1 M HCl, 1 ml; autoclaved 4-day old *Sorangium cellulosum* broth, 100 ml. A sample of cells was removed from the plates with a sterile loop and inoculated into 50 ml of G51t medium in a 250 ml Erlenmeyer flask. G51t consists of 0.5% starch (Cerestar), 0.2% tryptone, 0.1% yeast extract, 0.05% CaCl₂, 0.05% MgSO₄·7H₂O, 1.2% 4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid (HEPES), 0.2% glucose, pH 7.6. The flasks were shaken at 30°C, 160 rpm until a dense orange bacterial growth was obtained (ca. 5-7 d.). The cells were pelleted by centrifugation at 6,000 x g and used immediately or stored frozen at -20°C.

30

35

40

The protocol used for isolating chromosomal DNA from bacteria using hexadecyltrimethylammonium bromide (CTAB) has been described previously (Ausubel et al., Current Protocols in Molecular Biology, John Wiley and Sons, New York, 1990). The precipitated DNA was recovered with a bent Pasteur pipette, washed

50

55

5

8

10

with 70% and 95% ethanol, air-dried, and resuspended in 0.5 ml TE buffer (0.01 M Tris-HCl, 0.001 M ethylenediaminetetraacetic acid [EDTA], pH 8.0).

15

20

- 5 b. Alternatively, genomic DNA was isolated from *S. cellulosum* cells cultured as described in section A.1 using the Midi Qiagen Blood & Cell Culture DNA purification Kit (Qiagen, Hilden, Germany) following the Qiagen Genomic DNA Handbook protocol for bacterial DNA isolation (1997, Qiagen, Hilden, Germany, p. 29 ff.). In order to obtain high molecular weight chromosomal DNA the precipitated DNA was recovered with a bent pasteur pipette as described in section A.1.

25

15 2. Isolation of plasmid DNA

30

35

40

- a. pFD666: pFD666 is a bifunctional *E. coli*-*Streptomyces* cosmid cloning vector (see Denis and Brzezinski, Gene 111, 115-118 [1992]). To maintain stability of large inserts, it is present in low-medium copy number when replicated in *E. coli*. For this reason, isolation of sufficient pure DNA to carry out cloning experiments was difficult using commercial kits with standard protocols. A modified procedure was therefore used to obtain pFD666 DNA. A 10 ml culture of DH10B(pFD666) was grown for 16-20 hr at 37°C in LB (1% tryptone, 0.5% yeast extract, 0.5% NaCl, pH 7.0) medium containing 50 µg/ml kanamycin sulfate. Fifty ml of LB + kanamycin was inoculated to a starting OD₆₀₀ of ca. 0.25 and shaken at 300 rpm, 37°C, until the OD₆₀₀ reached ca. 0.6. Five hundred ml of LB + kanamycin medium in a 2 l flask was inoculated with 25 ml of this culture and incubated

45

30

50

55

5

9

10

15

20

25

30

35

40

45

50

55

under the same conditions for 2.5 hr. Chloramphenicol (2.5 ml of a 34 mg/ml solution in 100% EtOH) was added and the incubation continued for an additional 16-20 hr. (The previous steps were performed according to Maniatis et al. Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989.) Cells were pelleted for 10 min, 16,000 x g . They were resuspended in 9 ml of 50 mM glucose/25 mM Tris-HCl (pH 8.0)/10 mM EDTA and transferred to a 50 ml disposable centrifuge tube. One ml of a freshly-prepared 10 mg/ml lysozyme solution in 10 mM Tris-HCl, pH 8.0 was added and the cell suspension incubated in a 37°C water bath for 10 min. Twenty ml of a freshly-prepared 0.2 NaOH/1% sodium dodecyl sulfate (SDS) solution was added and the tube inverted gently 5-7 times to mix the contents. After 5 min at room temperature, 15 ml of 5 M potassium acetate (pH 4.8) was added and the tube inverted sharply 3-4 times. The tube was centrifuged at 6,000 x g for 10 min at 4°C and the supernatant poured through 2 layers of sterile cheese cloth into a fresh 50 ml disposable tube. Isopropanol to a final concentration of 0.6% was added and the contents of the tube mixed several times. The precipitated nucleic acid was centrifuged at 6,000 x g for 10 min at 4°C. The pellet was washed with 70% EtOH and any excess EtOH was aspirated from the pellet, which was allowed to air dry for 5 min. It was resuspended in 5 ml of 50 mM 3-(N-Morpholino)propanesulfonic acid (MOPS)/750 mM NaCl, pH 7.0 and added to an equilibrated QIAfilter Midi column (Qiagen, Chatsworth, CA). The manufacturer's protocol for washing and eluting the plasmid DNA was followed.

5

10

b. SuperCos: SuperCos plasmid DNA was purchased from Stratagene (La Jolla, CA).

10

3. Preparation of ca. 38-47 kb *Sau3A1* fragments of *S. cellul-
sum* chromosomal DNA

15

a. *S. cellulorum* chromosomal DNA prepared as described in section A.1.a was partially cleaved with restriction endonuclease *Sau3A1* in a 1000 μ l reaction volume consisting of 50 μ g chromosomal DNA, 5 units enzyme (Promega, Madison, WI), 0.006 M Tris-HCl, 0.006 M MgCl₂, 0.10 M NaCl, and 0.001 M dithiothreitol (pH 7.5) for 5 min at 37°C. The reaction mixture was extracted once with an equal volume of 1:1 phenol:chloroform. After centrifugation, the upper aqueous phase was saved, to which 0.1 vol. of 3 M sodium acetate and 0.6 vol. isopropanol was added. DNA was pelleted by centrifugation for 5 min at 16,000 x g in a microfuge and washed once with 0.5 ml 70% EtOH. After drying in a SpeedVac (Savant Instruments, Farmingdale, NY) for 5 min, the pellet was resuspended in 0.1 ml TE buffer. The DNA was layered on top of a 12 ml 10-40% sucrose gradient prepared in TE buffer and centrifuged at 113,600 x g for 16 hr, 10°C using a Beckman SW40Ti rotor (Beckman Instruments, Palo Alto, CA). Five hundred μ l aliquots of the gradient were removed using a pipetor beginning at the top of the tube. Samples (5 μ l) of the fractions were analyzed by electrophoresis through a 0.5% agarose gel in TAE buffer (0.04 M Trizma base, 0.02 M acetic acid, and 0.001 M EDTA, pH 8.3) containing 0.5 μ g/ml ethidium bromide for 6 hr at 100 V. Fractions containing DNA fragments of ca. 40-45 kb were identified by comparison to a high molecular weight DNA standard (Life Technologies,

20

10

25

15

30

35

40

45

50

55

5

10

15

Gaithersburg, MD). Sucrose was diluted from the corresponding 0.5 ml fraction by addition of 0.5 vol. TE. Subsequently, DNA was precipitated by addition of 0.1 vol. 3 M sodium acetate and 0.6 vol. isopropanol. DNA was pelleted by centrifugation at 16,000 x g for 10 min in a microfuge. DNA was washed with 0.5 ml 70% EtOH and dried in a SpeedVac with moderate heat for 10 min. Finally, the DNA was resuspended in distilled H₂O at a concentration of 0.5 mg/ml.

20

25

30

35

40

45

10 b. Alternatively, 10 µg of *S. cellulosum* chromosomal DNA prepared as described in A.1.b was treated with 0.3 U Sau3A1 (New England Biolabs, Beverly, MA) for 1 h at 37°C in 400 µl of the supplier's recommended reaction buffer. Formation of DNA fragments of about 40 kb in size was checked by comparison of the motility behavior with high molecular weight DNA standards after a 0.3% agarose gel electrophoresis. An equal volume of phenol:chloroform (1:1) was added, mixed and centrifuged. The upper aqueous phase was recovered and 0.1 vol. of 3 M sodium acetate and 0.6 vol. of isopropanol were added. After centrifugation, the precipitated DNA was washed twice with 0.5 ml 70% ice cold ethanol and finally air-dried. The DNA fragments were resuspended in 100 µl shrimp alkaline phosphatase reaction buffer and dephosphorylated for 150 min. at 37°C using 2 U shrimp alkaline phosphatase (Amersham Life Science, Cleveland, OH). A phenol:chloroform extraction followed as described above. Finally, the DNA was precipitated by addition of 0.1 vol. 3 M sodium acetate and 0.6 vol. isopropanol, dried, and dissolved in TE buffer.

30

50

55

5

12

4. Preparation of cosmid libraries

10

5

15

20

25

30

35

40

45

50

55

a. Using pFD666: Vector pFD666 was cleaved with restriction endonuclease BamHI in a 0.02 ml reaction volume consisting of 2 μ g plasmid DNA, 10 units of BamHI (Promega), 0.006 M Tris-HCl, 0.006 M MgCl₂, 0.05 M NaCl, and 0.001 M dithiothreitol (pH 7.5) for 90 min at 37°C. Five μ l of 10x alkaline phosphatase buffer (0.5 M Tris-HCl [pH 9.3], 0.01 M MgCl₂, 0.001 M ZnCl₂, 0.01 M spermidine) was added to the reaction followed by alkaline phosphatase (0.01 units/pmol ends; Promega) and distilled H₂O to a final volume of 0.05 ml. The sample was incubated for 30 min at 37°C and a second aliquot of phosphatase was added. After a further 30 min at 37°C, 0.3 ml of stop buffer (0.01 M Tris-HCl [pH 7.5], 0.001 M EDTA, 0.2 M NaCl, 0.5% SDS) and 0.35 ml of 1:1 phenol; CHCl₃ was added to the reaction. The sample was mixed gently several times by inversion and centrifuged at 16,000 x g for 3 min to separate the phases. The aqueous layer was removed to a new microfuge tube. 0.1 vol. 3 M sodium acetate and 2 vol. 100% EtOH were added and the precipitated DNA pelleted by centrifugation at 16,000 x g for 10 min. Liquid was removed by aspiration and the pellet washed once with 0.5 ml 70% EtOH. The DNA was dried in a SpeedVac and resuspended in TE buffer to 0.5 mg/ml.

Digested, phosphatase-treated pFD666 was ligated to the partially-cleaved chromosomal DNA (see sections A.3.a and B.1.a) in a 0.005 ml reaction consisting of 1 μ g pFD666, 1 μ g *S. cellulosum* DNA, 0.03 M Tris-HCl (pH 7.8), 0.01 M MgCl₂, 0.01 M dithiothreitol, and 0.0005 M adenosine-5'-triphosphate and 1.5 Weiss units of T4 DNA ligase (Promega). The reaction was carried out at room temperature for 2 hr. The entire reaction

5

13

10

15

20

25

30

35

40

45

50

55

mix was packaged into bacteriophage λ in vitro using Packagene extracts (Promega) according to the manufacturer's directions. The entire packaging reaction (0.5 ml) was diluted with 4.5 ml SM buffer (per liter: 5.8 g NaCl, 2 g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 1 M Tris-HCl [pH 7.5], 5 ml 2% gelatin solution). Transfection was performed by adding 10 ml of an overnight culture of *E. coli* DH5 α that had been grown in LB medium with 0.01 M MgSO_4 and 0.2% maltose to the diluted phage and incubating at 37°C for 20 min. 0.8 ml of LB was added and the cells shaken at 225 rpm for 1 hr at 37°C. Cells were pelleted, resuspended in LB, and spread onto a 150 mm LB + kanamycin agar plate. After 3 d. at 30°C, the colonies were harvested by picking ca. 800 colonies into 2.0 ml LB + kanamycin medium containing 20% glycerol, freezing on dry ice, and storing at -70°C. In addition, six kanamycin-resistant colonies were inoculated into 2 ml LB + kanamycin liquid medium and incubated at 37°C, 250 rpm, for 18-24 hr. Cosmid DNA was prepared using a standard alkaline lysis procedure starting with 1.5 ml of the culture. DNA was digested with restriction endonuclease PstI and samples electrophoresed on a 0.8% TAE agarose gel for 1.5 hr at 100 V. A unique restriction pattern was noted in each sample and the total size of the insert was calculated to be between 40 and 45 kilobases.

b. Using SuperCos: 30 μg of vector SuperCos was digested with XbaI (New England Biolabs, Beverly, MA) for 210 min at 37°C in 100 μl of the recommended reaction buffer. Ten μl sodium acetate and 60 μl isopropanol was added before the solution was centrifuged for 30 min at 16,000 \times g. The precipitated DNA was washed twice with 500 μl ice cold 70% ethanol. The vector DNA was precipitated and air-dried, dissolved in 135 μl shrimp al-

5

10

15

20

25

30

35

40

45

50

55

kaline phosphatase reaction buffer and treated with 2.5 U shrimp alkaline phosphatase for 150 min. After heat inactivation of the enzyme at 75°C for 20 min, a phenol:chloroform extraction was performed as described in section 1. c. The DNA, resuspended in 100 μ l BamHI restriction buffer was hydrolyzed with 15 U BamHI (New England Biolabs, Beverly, MA) for 180 min. A phenol:chloroform extraction followed (see section A.3). The SuperCos DNA was precipitated by addition of 0.1 vol 3 M sodium acetate and 0.6 vol isopropanol, centrifuged at 16,000 x g, and resuspended in 50 μ l TE buffer.

Four μ g of digested vector DNA was ligated with 10 μ g partially hydrolyzed genomic DNA from *S. cellulosum* (as described in section A.3.b) in a final volume of 20 μ l using 2 U T4 DNA ligase and the appropriate reaction buffer (Gibco BRL, Eggenstein, Germany). The reaction was carried out at 16°C overnight. The reaction mixture was packaged into phage particles using the Gigapack III XL packaging extract kit (Stratagene) according to the manufacturer's protocol. Treatment of packaging reaction mixture and transfection of *E. coli* SURE (Stratagene) was performed as described in 4.a. Transfected cells were concentrated by centrifugation, resuspended in fresh LB medium and distributed on LB agar plates containing 50 μ g/ml kanamycin. The plates were incubated overnight at 30°C. 1600 recombinant clones were transferred into 96 well microtiter plates filled with 80 μ l LB medium containing 50 μ g/ml kanamycin per well and propagated overnight at 30°C. The following day the microtiter plates were used to inoculate a second set of microtiter plates in order to obtain a duplicate of the recombinant clones. Each well of the original set of microtiter plates was supplemented with 80 μ l 50 % glycerol and the entire plate stored at -70°C.

20 randomly chosen transformants were inoculated into 3 ml LB medium with 50 $\mu\text{g}/\text{ml}^{-1}$ kanamycin and incubated over night at 37°C in order to isolate plasmid DNA using the Qiagen plasmid extraction kit (Qiagen, Hilden, Germany). Restriction fragment analysis of the recombinant cosmids using the restriction endonucleases PstI and BglII indicated that the cosmids contained inserts of approximately 35 to 42 kb in size.

B. Construction of a *S. cellulosum* plasmid library

1. Preparation of 8-12 kb fragments of *S. cellulosum* chromosomal DNA.

S. cellulosum chromosomal DNA prepared as described in section A.1.a was partially cleaved with restriction endonuclease Sau3AI in a 100 μL reaction volume consisting of 5 μg chromosomal DNA, 5 units enzyme (Promega, Madison, WI), 0.006 M Tris-HCl, 0.006 M MgCl_2 , 0.10 M NaCl, and 0.001 M dithiothreitol (pH 7.5) for 4 min at 37°C. The digested DNA was electrophoresed through a 11 x 14 cm 0.8% TAE-agarose gel for 18 hr at 17 V. Fragments of 8-12 kb were cut from the gel and purified using the QIAquick Gel Extraction Kit using the manufacturer's protocol (Qiagen).

2. Preparation of the plasmid library

Plasmid pZero2.1 (Invitrogen, Carlsbad, CA) was cleaved with restriction endonuclease BamHI in a 0.02 ml reaction volume consisting of 1 μg plasmid DNA, 10 units of BamHI (Promega), 0.336 M Tris-HCl, 0.006 M MgCl_2 , 0.05 M NaCl, and 0.001 M di-

5

16

10

15

20

25

30

35

40

45

50

55

thiothreitol (pH 7.5) for 20 min at 37°C. 0.08 ml of dH₂O and 0.1 ml of 1:1 phenol:CHCl₃ was added. The sample was briefly vortexed and centrifuged at 16,000 x g for 2 min. The aqueous layer was removed to a new microfuge tube. 0.1 vol. 3 M sodium acetate and 2 vol. 100% EtOH were added and the precipitated DNA pelleted by centrifugation at 16,000 x g for 10 min. Liquid was removed by aspiration and the pellet washed once with 0.5 ml 70% EtOH. The DNA was dried in a SpeedVac and resuspended in TE buffer to 0.004 µg/ml. Digested pZero2.1 was ligated to the partially-cleaved chromosomal DNA in a 0.01 ml reaction consisting of 0.004 µg pZero2.1, 0.05 µg *S. cellulosum* DNA, 0.03 M Tris-HCl (pH 7.8), 0.01 M MgCl₂, 0.01 M dithiothreitol, and 0.0005 M adenosine-5'-triphosphate and 1.5 Weiss units of T4 DNA ligase (Promega). The reaction was carried out at room temperature for 2 hr. 0.015 ml dH₂O and 0.25 ml of 1-butanol were added, the sample vortexed briefly, and centrifuged at 16,000 x g for 10 min. Liquid was aspirated away from the pellet and the sample dried in a SpeedVac for 5 min. The ligated DNA was resuspended in 0.005 ml dH₂O and mixed with 0.04 ml of electrocompetent *Escherichia coli* DH10B cells (GIBCO/BRL, Gaithersburg, MD). The sample was placed into a pre-chilled 0.2 mm-gap electroporation cuvette and transformed into the bacteria by electroporation using a BioRad Gene Pulser II unit (BioRad, Hercules, CA) at 25 µF and 200 Ω. 0.96 ml SOC medium (0.5% yeast extract, 2% tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 20 mM MgSO₄, 20 mM glucose) was mixed with the cells and transferred to a 1.5 ml microfuge tube. The sample was incubated at 37°C, 225 rpm, for 1 hr. Aliquots of the cells were spread onto an LB agar + kanamycin and incubated at 37°C for 20 hr to estimate the number of transformants obtained. Six kana-

mycin resistant colonies were confirmed to contain an insert of the expected size as described in section A.4.a.

C. Identification of cosmids possessing polyketide synthase genes

1. Colony blot hybridizations using cosmid library in pFD666:

A 20 x 20 cm sheet of Duralon UV membrane (Stratagene) was placed on top of a 24.5 x 24.5 cm square bioassay dish containing 250 ml LB agar - kanamycin. An aliquot of the frozen cosmid library in 1 ml LB medium was spread on the filter. The plate was incubated at 37°C for 24 hr. Colonies were replicated onto two fresh filters which were placed onto LB + kanamycin agar medium and incubated at 28°C for 18 hr. Lysis of cells and neutralization of released DNA was performed according to directions that were provided with the filters. The DNA was crosslinked to the filters using a UV Stratalinker 2400 unit (Stratagene) in the auto crosslink mode. Cell debris was removed by placing the filters in a container with a solution of 3 X SSC (20 X SSC contains, per liter, 173.5 g NaCl, 88.2 g sodium citrate, pH adjusted to 7.0 with 10 N NaOH), 0.1% SDS and rubbing the lysed colonies with a Kimwipe. The filters were then incubated at least 3 hr with the same wash solution for at least 3 hr at 65°C. The plasmid library was treated similarly except cells were spread onto a 137 mm circular Duralon UV membrane placed on top of a 150 mm petri dish containing 80 ml LB agar + kanamycin.

For hybridizations, a probe consisting of a 650-base pair (bp) polymerase chain (PCR) fragment representing a portion of a *S. cellulosum* polyketide synthase gene was used. The fragment

was amplified using primers to consensus regions of Type I (macrolide) polyketide synthase (PKS) genes (Swan et al., Mol. Gen. Genetics 242, 358-362 [1994]). A series of sense and anti-sense oligonucleotides were prepared for PCR studies as indicated in the following table 2:

Table 2

Oligo-nucleotide	I. DNA sequence (5' → 3')	Corresponding amino acid sequence
120 (sense)	CGGT (C/G) AAGTC (C/G) AACATCGG	KSNIGHT
121 (anti-sense)	GC (A/G) ATCTC (A/G) CCCTGCGA (A/G) TG	HSQGEIA
122 (sense)	GT (C/G) GACAC (C/G) GC (C/G) TGCTC (C/G)	VDTACSS
123 (sense)	GG (C/G) AC (C/G) AACGC (C/G) CACGT (C/G) A T	GTNAHVI
124 (anti-sense)	CCCTG (C/G) CC (C/G) GGGAA (C/G) ACGAA	FVFPGQG

10

The selection of C or G where necessary in the third position of a codon reflects the very high overall G + C content of *S. cellulosum* (ca. 70%). Conditions for PCR were as follows: 0.01 M Tris-HCl (pH 9.0), 0.05 M KCl, 0.003 M MgCl₂, 0.1% Triton X-100, 200 μM of each primer, 2.5 U Taq DNA polymerase (Promega), 5.0% dimethyl sulfoxide (Sigma), and 0.01 μg of *S. cellulosum* chromosomal DNA in a 0.25 ml reaction volume. Reactions were

5

19

10

15

20

carried out in a Perkin-Elmer Model 480 Thermocycler (Perkin-Elmer Corporation, Foster City, CA) under the following conditions: 94°C, 1 min; 50°C, 1 min, 72°C, 1.5 min for a total of 30 cycles. Each possible combination of sense and anti-sense primers were tried. A 650-bp and 350-bp fragment was amplified using cligos 120 + 124 and 123 +124, respectively. The sequence of the fragments were determined using the ALFexpress AutoRead kit to fluorescently label the DNA, which was analyzed on an ALFexpress sequencing apparatus (Pharmacia). The data indicated both PCR fragments possessed significant homology to polyketide synthase genes of Type I antibiotics. The 650-bp fragment was chosen for hybridization experiments.

25

30

35

40

The fragment was labeled with ³²P-dCTP using the NEBlot kit (New England Biolabs, Beverly, MA) and purified on a Bio-Spin 6 column (BioRad, Hercules, CA.). Duplicate blots were pre-hybridized in 3 X SSC (1 X SSC contains 0.15 M sodium chloride and 0.015 M sodium citrate, pH 7.0), 4 X Denhardt's solution (100 X is 2% Ficoll [Type 400], 2% polyvinylpyrrolidone, and 2% bovine serum albumin [Fraction V]), and 100 µg/ml sheared, denatured salmon sperm DNA; all reagents purchased from Sigma Chemicals, St. Louis. The labeled DNA was heated in a boiling water bath for 5 min to denature the strands, cooled on ice, and added to the pre-hybridization solution. The filters were incubated for at least 18 hr in a roller bottle hybridization oven. They were transferred to new bottle, then washed two times in 2 X SSC, 0.1% SDS at 70°C for 30 min (moderate stringency). The membranes were placed on Whatman 3MM paper to remove excess liquid, covered with Saran Wrap, and exposed to autoradiography film (Kodak X-OMAT LS) with two intensifying

45

50

55

5

20

5 screens. The cassette was placed at -70°C and developed at appropriate intervals.

10

Approximately 100 colonies were seen to have hybridized on the duplicate filters. Fourteen of these were isolated from the master plate and grown in 4 ml LB + kanamycin medium for 20-24 hr, 37°C, 250 rpm. Plasmid DNA was prepared using the standard alkaline lysis method and digested with restriction endonuclease PstI. The digested DNA was electrophoresed on a 0.8% agarose gel in TAE for 3 hr at 100 V. Fragments were transferred to Duralon UV using the VacuGene XL vacuum blotting unit (Pharmacia) and the recommended alkaline denaturation protocol. Hybridization with radioactively-labeled PCR fragment and washing were carried out as described above. Two prominent types of cosmids were observed; one contained PstI fragments of ca. 7.0, 5.0, and 1.1 kb (pEPOcos5 and pEPOcos7) that hybridized to the probe; the other type had fragments of ca. 6.0 and 3.6 kb (pEPOcos8 and pEPOcos13) which were homologous to the probe. Restriction analysis confirmed that cosmids showing identical hybridization patterns had identical or overlapping inserts. PCR reactions using primers representing consensus sequences of Type I PKS genes were performed using the isolated cosmid DNA as template under conditions described above, except ca. 0.01 µg of cosmid DNA was included as template. Cosmids pEPOcos6 and pEPOcos8 amplified the 650-bp fragment seen when oligonucleotides 120 + 124 were used, while pEPOcos8 and pEPOcos13 supported amplification of an 1100-bp PCR fragment with oligos 122 and 124. The latter fragment was sequenced and confirmed to possess strong similarity to Type I PKS genes. These data confirm that the recombinant cosmids are related to each other and that all contain PKS-like genes.

25

30

35

40

45

50

55

5

21

10

2. Colony blot hybridizations of plasmid library in pZero2.1:

15

20

25

A 137-mm circle of Duralon UV membrane was placed on top of a 150-mm containing 75 ml LB agar + kanamycin. An aliquot of the plasmid library (representing ca. 2,000 recombinant colonies) in 0.5 ml LB medium was spread on the filter. The plate was incubated at 37°C for 20 hr. Colonies were replicated onto two fresh filters which were placed onto LB + kanamycin agar medium and incubated at 37°C for 6 hr. The filters were processed for hybridization as described in Section C.1. Out of 8 positive colonies detected, one contained a plasmid with a DNA region not encoded by either pEPOcos6 or pEPOcos8. This plasmid, called Sau4, was characterized in more detail.

3. Colony blot hybridizations of cosmid library in SuperCos:

30

35

40

45

50

The recombinant *E. coli* clones from the microtiter plates (see section 4. b) were used to produce two identical sets of hybridization filters in order to identify cosmids carrying PKS and PS genes. The recombinant clones were spotted onto 2 sets of 22 x 22 cm LB agar plates containing 50 µg/ml kanamycin. Each plate contained 384 clones therefore representing 4 microtiter plates. The clones were incubated at 30°C overnight. After pre-cooling for approximately 3 h at 4°C, 20 x 20 cm Hybond N⁺ Nylon membranes (Amersham, Braunschweig, Germany) were placed onto the agar surfaces. After 2 min. the membranes were removed and placed for 15 min. on Whatman 3 MM paper (Whatman paper Ltd., Maidstone, England) soaked with denaturation solution (0.5 N NaOH, 1.5 M NaCl) before they were transferred onto Whatman 3 MM paper saturated with neutralization solution (1 M Tris-HCl, pH 7.5, 1.5 M NaCl). Subsequently the membranes were

55

5

22

10

15

placed onto Whatman 3 MM paper soaked with 2 X SSC (0.3 M NaCl, 0.03 M sodium citrate, pH 7.2) for 10 min. The membranes were baked for 40 min at 85°C. Then, each membrane was overlaid with 5 ml Proteinase K solution (2 mg/ml Proteinase K in 2 x SSC) and incubated at 37 °C for 90 min. Finally, cell debris was removed by wiping the membranes with a Kimwipe pre-wetted with 2 X SSC.

20

25

30

As we were seeking in particular to identify biosynthetic pathways containing both PKS and PS genes, the following hybridization strategy was taken: The screening was initially focused on ketosynthase domains from type I PKSs and on the adenylation domain from PSs. Target-specific primers were used to amplify DNA fragments of the corresponding genes from chromosomal DNA of *S. cellulosum* by PCR. The fragments obtained were then cloned, sequenced and the deduced amino acid sequence compared to known ketosynthase and adenylation domains of PKS and PS, respectively. In a second step these PCR fragments were used as gene probes to detect recombinant cosmids of the *S. cellulosum* cosmid library.

35

40

Oligonucleotides based on conserved amino acid sequences of ketosynthase domains from various type I PKS were optimized for myxobacterial DNA by comparison to a known myxobacterial biosynthetic gene cluster (Schupp et al., J. Bacteriol. 177, 3673-3679 [1995]) resulting in primer

KS1Up (5'-

C/A)GIGA(A/G)GCI(A/C/T)(A/T)I(C/G)(C/A)IATGGA(C/T)CCICA(A/G)CAI(A/C)G-3') and

45

KSE1 (5'-GG(A/G)TCICCA(A/G)I(G/C)(T/A)IGTICGIGTICC(A/G)TG-3').

PCR-primers TGD (5'-

T(A/T)(C/T)CGIACIGGIGA(C/T)(C/T)(G/T)IG(G/T)ICG-3') and

50

55

5

23

LGG (5'-

10

A(A/T)IGA(A/G)(G/T)(G/C)ICCI(A/G)(A/G)(G/C)I(A/C)(A/G)AA(A/G)
)AA-3')

15

20

25

30

35

40

45

50

55

directed to genes encoding adenylation modules have been described by Turgay et al. (Pept. Res. 7, 238-241 [1994]). PCR reaction mixtures with a final volume of 25 μ l contained 0.1 μ g template DNA, 0.2 U Taq DNA-polymerase (Gibco BRL, Eggenstein, Germany), 5 μ mol dNTP, 5% dimethyl sulfoxide (Sigma), 1.5 mM $MgCl_2$, 25 pmol of each primer and the appropriate reaction buffer supplied by Gibco BRL. Chromosomal DNA of *S. celluloseum* was used as template. Additionally, chromosomal DNA of *Myxococcus fulvus* was used with PS primers. Reactions were carried out in an Eppendorf Mastercycler Gradient (Eppendorf, Germany) using the following conditions: denaturation 30 s at 97°C, annealing 30 s at 55°C, extension 60 s at 72°C for a total of 30 cycles. The formation of ca. 700 bp fragments using the KS primers and of ca. 350 bp fragments with the PS primers were confirmed by 0.8% agarose gel electrophoresis. Fragments of independent PCR reactions were ligated into vector pCR2.1TOPO using the TOPO TA Cloning kit (Invitrogen, Leek, The Netherlands) according to the manufacturer's protocol and transformed into *E. coli* XL1-Blue. Sequencing of the resulting plasmids and analysis of the deduced amino acid sequence revealed three different KS fragments, designated pM008.4, pM008.6, pM008.7, one PS fragment (pAPs1) corresponding to *S. celluloseum* and one PS fragment (pDPs1) obtained with chromosomal DNA of *M. fulvus*. The PCR fragments were re-isolated by digestion with *Eco*RI from the plasmids pM008.4, pM008.6, and pM008.7, labeled, pooled and used as gene probes in hybridization experiments as described

5

24

below. The same procedure was performed with the PS fragments of pAPs1 and pDPs1.

10

Hybridization with PKS and PS specific DNA probes (see above) was carried out using the DIG nonradioactive labeling and detection kit (Boehringer Mannheim, Germany) and performed according to the supplier's manual using buffer containing 50% formamide. The membranes were hybridized in plastic bags containing approx. 10 ml of hybridization solution at 39°C overnight. Unspecific binding of probes was removed by 2 wash steps with 2 x SSC, 0.1% SDS at room temperature for 20 min. and one stringent wash step with 0.5 x SSC, 0.1% SDS at 60°C for 20 min. Detection of hybridizing DNA fragments was performed with the above mentioned system according to the manufacturer's protocol using CSPD as chemiluminescent substrate. The signals were recorded by exposure of the treated membrane to Hyperfilm ECL (Amersham Life Science, Little Chalfont, England) which was developed in appropriate time intervals.

20

10

25

15

30

71 signals were detected with the PKS specific gene probe. On the duplicate filters 35 signals were obtained with the PS specific gene probe of which 7 were already known from the PKS hybridization experiment. These recombinant cosmids harbored PKS- and PS-encoding genes. In order to corroborate these results PCR experiments were performed with DNA of the 7 recombinant cosmids as template and PKS (KS1Up, KSD1) and PS specific primers (TGD, LGG) generating fragments of the expected size of approx. 700 bp and 350 bp, respectively (primers and reaction conditions see above).

35

40

25

A comparison of the restriction fragment patterns of the DNA from the 7 recombinant cosmids carrying PKS and PS genes digested by BamHI facilitated an arrangement of the cosmids in

45

30

50

55

5

25

10

3 groups. They were represented by cosmids designated A2 and A5. The remaining group was represented by pEPOcos6. Therefore, A2 and A5 represented good candidates for further DNA sequence analysis because they carry both PKS and PS genes.

5

15

D. Random "shotgun" sequencing of recombinant cosmids and plasmids

20

1. Library construction

10

25

30

a. pEPOcos6, pEPOcos8, A5, and Sau4: pEPOcos6 and pEPOcos7 were sequenced to completion, and contiguous sequence data and analysis for these overlapping cosmids is presented below for the "cos6 region" (cf. claims 7 and 9). Sequencing of cosmid A5, pEPOcos8 and plasmid Sau4 was taken to the point of large contiguous sequences (contigs) representing the *S. cellulorum* insert; sequence and analysis presented below (cf. claims 10 to 15).

35

40

45

Randomly sheared libraries were constructed for cosmids and plasmids of interest using a protocol similar to that of of Fleischmann et al., 1993 (Science 269, 496) and modified in Fraser et al., 1995 (Science 370, 397). Briefly, Qiagen-column purified cosmid DNA (~10 µg) was sheared to a size of approximately 2 kb and the DNA end-repaired using BAL31 nuclease. The DNA was gel-purified after electrophoresis through a 0.75% low-melting temperature agarose gel containing 0.5 µg/ml ethidium bromide in 1X TAE buffer run at 80 V for 2 hours. The volume of the low-melt agarose gel slice was estimated by adding the gel slice to a microfuge tube and weighing, then 0.1 vol. of 3 M sodium acetate (pH 7) was added and the agarose incubated at 60°C. The temperature was equilibrated to 37°C, and DNA ex-

50

55

5

26

10

15

20

25

15

30

35

40

25

tracted twice using an equal volume of buffered phenol (Life Technologies). The aqueous phase was transferred and extracted once with an equal volume of chloroform, then ethanol precipitated by the addition of 2 vol. cold 100% ethanol. DNA was concentrated by spinning at 16,000 x g in a microcentrifuge. The DNA pellet was washed with 1 ml 70% ethanol and resuspended in 100 μ l of 0.1X TE. The DNA was ligated to SmaI-digested, phosphatase-treated pUC18 vector (Pharmacia), and single insert recombinants isolated by gel-purification of the band containing vector plus a single insert, followed by T4 polymerase polishing, and a final intramolecular ligation of the vector-plus-single-insert DNA. This final ligation represents a library of highly random ca. 2 kb fragments that was used for shotgun sequencing of the ca. 40 kb cosmids or ca. 10 kb plasmids.

b. Cosmid A2: Cosmid DNA with inserts of *S. cellulorum* was isolated by an alkaline lysis procedure and purified with Macherey Nagel columns (Machery und Nagel GmbH und CoKG, Düren, Germany) using manufacturer's recommendation. Purified Cosmid DNA was sonicated, end-repaired using T4 DNA Polymerase (Boehringer Mannheim, Germany). After gel-purification fragments of a size of approximately 2 kb were ligated into SmaI-digested, phosphatase-treated pTZ18R vector (Pharmacia). The ligation represents a library of highly random ca. 2 kb fragments that was used for shotgun sequencing of the ca. 40 kb cosmid.

2. Sequencing and assembly

45

a. pEPOcos6, pEPOcos8, Sau4, and A5: DNA (1 μ l of 100 μ l total in the library) was transformed into *E. coli* by electroporation (20 μ l of Electromax DH10B cells from Life Technolo-

50

55

5

27

10

15

20

25

gies) and cells spread onto LB plates containing 50 µg/ml ampicillin. After growth overnight at 37 °C, transformants (ca. 300-3000 CFU total) were transferred to 96-well growth blocks and shaken overnight at 37°C in 1.3 ml LB medium with 50 µg/ml ampicillin. Templates were prepared from these cells by an alkaline lysis procedure (Qiagen QiaQuick Turbo Prep) to yield purified, double-stranded plasmid DNA. Cycle-sequencing of the plasmid templates was performed using universal forward and reverse primers and BigDye Terminator sequencing kits (Applied Biosystems), using the manufacturer's recommendations, then resolved using an ABI377 automated sequencer. Sequences were edited using Phred, then assembled into larger contiguous sequences using Phrap (Phil Green, University of Washington, St. Louis, MO).

15

30

35

40

45

b. Cosmid A2: DNA (1 µl of 20 µl total in the ligation) was transformed into *E. coli* DH10B by electroporation and cells were spread onto LB agar medium containing 50 mg/ml ampicillin. After growth for 18 hr at 37°C, transformants were transferred to 96-well growth blocks and shaken overnight at 37°C in 1.3 ml 2x YT medium with 50 mg/ml ampicillin. Templates were prepared from these cells by an alkaline lysis procedure (Qiagen QiaQuick Turbo Prep) to yield purified, double-stranded plasmid DNA. Cycle-sequencing of the plasmid templates was performed using universal forward and reverse primers and Big Dye Terminator sequencing kits (PE Biosystems) or Thermo Sequenase fluorescent labelled primer cycle sequencing kit (Amersham Pharmacia Biotech) using the manufacturer's protocols. In the shotgun phase of a cosmid, identical amounts of samples were sequenced either by dye-primer or dye-terminator chemistries (Pharmacia, PE Biosystems). Data were collected using Licor and ABI 377

50

55

5

28

10

automated sequencers and assembled with the GAP4 program (Bonfield, Smith, Staden, Nucl. Acids Res. 23, 4992-4999 [1995]). Gaps were closed using custom made primers (MWG-Biotech) on plasmid templates or PCR products in combination with dye-terminators.

15

E. Bioinformatic Methods

20

1. Open reading frame (ORF) identification

10

ORFs were identified in the pEPOcos6 region using the OMIGA 1.1.2 (GCG 0.4D) program from Oxford Molecular Limited. Default values were used (Standard genetic code, all ORFs over 50 bases) to generate ORFs; analysis of these results lead to the list of 14 highest quality ORFs as defined in claim 9. Other ORFs, genes, or genetic elements may be found in the pEPOcos6 insert that have not yet been annotated. In addition to hand-editing of the OMIGA-generated data, the MAGPIE automated genome analysis tool:

25

30

(<http://genomes.rockefeller.edu/magpie/magpie.html>)

35

20

was used to identify genes for all the sequenced cosmids and plasmids. ORFs identified in this manner are presented as both nucleotide and peptide files below.

40

For cosmids A2 and A5, ORFs have been identified within the DNA sequences of A5 (contigs 10, 11, 12) and of A2 using the FramePlot analysis program from Ishikawa and Hotta (FEMS Microbiol. Lett., 174, 251-253 [1999] public available under <http://www.nih.go.jp/~jun/cgi-bin/frameplot.pl>) which is based on positional base preference in codons typical for organisms having genomes with a high G + C content (Bibb et al., Gene 30,

45

30

50

55

157-166 [1984])). Default parameters using ATG and GTG as start codons were used. The deduced amino acid sequence of predicted ORFs were compared with protein databases (GenBank, CDS translations, PDB, SwissProt, PIR, PRF) using BLASTP (Altschul et al., Nucleic Acids Res., 25, 3389-3402 [1997])). Additionally, high scoring amino acid sequences were analyzed using the Pfam program [<http://www.sanger.ac.uk/Software/Pfam/>], which identified specific domain structures of the submitted proteins (Bateman et al. Nucleic Acids Res., 27, 260-262 [1999])).

2. BLAST searches

BLASTP2 similarity searches were performed using the peptide files from the above ORF identification strategy as query sequences. Searches were performed using the in-house Bioinformatics BLASTP2 (Version: BLASTP 2.0a19MP-WashU) web page at the Bristol-Myers Squibb Pharmaceutical Research Institute (allows BlastN2, BlastP2, BlastX2, TblastN, and TblastX searches). In addition, peptide files generated by the MAGPIE analysis were automatically searched using a FASTA algorithm.

3. Best match and probable identification

Analysis of the BLASTP2 and FASTA output led to an assignment of a best match and probable function. The best match was usually the top scoring match, although sometimes another match was given because it was a more relevant homolog, or no match was found with a significance greater than $>e-4$. Probable function represents the best estimate of function given the initial analysis of the BLAST data and the published literature regarding the best match, and may not necessarily represent the true function of the gene product (hypothetical proteins are of un-

5

30

10

15

known function). A higher probability score indicates a higher likelihood that the probable function corresponds to that of the best match; e.g., the polyketide synthase matches are all above e-100, and given the very high significance scores are presumed to function as polyketide synthases (as are the high scoring peptide synthetases).

The following is a summary of the sequence data from the pEPOcos6 region, pEPOcos8, A5, Sau4 and A2.

20

10 a. Data from pEPOcos6 region:

25

Summary: A large PKS/PS cluster spanning multiple cosmids. An IS element (designated IS-Sc1 here) is found in the cluster - this may be a potential tool for genetic analysis of *Sorangium*.

15

30

Statistics: Sequence was assembled from over 2000 random sequences (forward and reverse reads of the ca. 2 kb cloned fragments derived).

20

35

47,713 nucleotides of contiguous sequence (no pFD666 vector included)

40

25

DNA sequence data are as defined in claim 7.

45

Note: pEPOcos6_ORF7 sequences (cf. claim 9): the predicted N-terminus of ORF7 shows 145 nucleotide overlap with ORF6.

30

Note: pEPOcos6_ORF8 sequences (cf. claim 9): >pEPOcos6_ORF8.seq ("ORF9_up" in Fig.2)

50

55

5

31

10

57.3% G+C

15

Table 3 shows ORF data summary. Note: pEPOcos6_ORF1.seq is truncated at its 5' end; correspondingly pEPOcos6_ORF1.pep is truncated at its N-terminus.

b. Data from pEPOcos8 region:

20

10 Summary: Two PKS genes found on a cosmid. A Tn1000 insertion is also found (occurred during *E. coli* propagation). No peptide synthetase genes were found; one P450 hydroxylase was identified.

25

30

15 Statistics: 1952 random sequence reads from the pEPOcos8 library were assembled using phrap, with 1024 of the sequences assembling into 57 contigs. 12 of these contigs were chosen (totaling 56,537 bp) which each contained >6 reads and consisted of about 1000 bp or more. The sequences of these 12 contigs and the associated ORFs are given below.

35

DNA sequence data from contigs are as defined in claim 10. Table 4 shows more data.

40

25 pEPOcos8 protein data are as defined in claim 11, i.e. for selected ORFs (polyketide synthase, peptide synthetases, or ORFs with high similarity to known genes).

45

50

55

5

32

c. Data from cosmid A5 insert:

10

Summary: A cluster of PKS and PS genes found on the cosmid. Other genes possibly involved in this secondary metabolite production include a downstream lipxygenase gene highly similar to eukaryotic orthologs.

15

20

Statistics: 880 random sequence reads from the A5 library were assembled using phrap, with 530 of the sequences assembling into 12 contigs. 3 of these contigs were chosen (totaling 41,556 bp) which each contained >100 reads and consisted of about 9000 bp or more. The sequences of these 3 contigs and the associated ORFs are given below.

25

DNA sequence data from contigs are as defined in claim 12. Table 5 shows more data.

30

Protein sequence data from selected A5 ORFs are as defined in claim 13.

20

35

d. Data from plasmid Sau4 insert:

40

Summary: Insert contains PKS genes on two large contigs - most similar to the soraphen PKS gene from Sorangium.

25

45

Statistics: 565 random sequence reads from the Sau4 library were assembled using phrap, with 84 of the sequences assembling into 18 contigs. 2 of these contigs were chosen (totaling 6596 bp) which each contained >10 reads and consisted of

50

55

about 1000 bp or more. The sequences of these 2 contigs and the associated ORFs are given below.

DNA sequence data from plasmid Sau4 contigs are as defined in claim 14. Table 6 shows more data.

Protein sequence data from selected plasmid Sau4 ORFs are as defined in claim 15.

e. Data from cosmid A2

Table 7 shows ORF data summary

F. Construction of suitable recombinant expression vectors

1. Expression in Myxobacteria

Heterologous expression of the ORFs shown in Figure 1 is performed by using a derivative of plasmid pSUP102 (Simon, R., Priefer, U., Pühler, A., Methods in Enzymology (1986), vol. 118, pp. 643-659). In this plasmid the gene for chloramphenicol resistance is changed for a cassette comprising the gene for streptomycin resistance and the promoter element of the Tn5 transposon. Short homologous genomic DNA segments from the host organism are ligated with the DNA sequences of Figure 1 and with efficient regulatory elements into, for example, the EcoRI restriction site of the vector. Following amplification of the vectors in *Escherichia coli* the DNA is transferred by electroporation of the host cells or by conjugation with *Escherichia coli* S17-I (Simon, R., Priefer, U., Pühler, A., Biotechnology (1983), vol. 1, pp. 784-791).

5

34

10

By means of the tetracycline or streptomycin resistance, respectively, mediated by the vector the host cells are checked for integration of recombinant plasmid DNA into the chromosome by homologous recombination.

5

15

2. Expression in Streptomyces cells

Heterologous expression of the ORFs shown in Figure 1 is performed by using bifunctional Streptomyces-Escherichia coli cosmids pKU206 and pOJ466.

20

10

3. Expression in Escherichia coli cells

Heterologous expression of the ORFs shown in Figure 1 is performed by using "bacterial artificial chromosomes", cosmids (for example Supercos, Stratagene GmbH, Heidelberg) and T7 expression systems (Stratagene GmbH, Heidelberg; New England Biolabs Schwalbach, FRG). Expression of recombinant enzymes occurs in Escherichia coli cells constitutively expressing phosphopantetheinyl transferase required for the formation of holoenzyme polyketide synthetases and polypeptide synthetases.

25

15

30

35

40

45

50

55

Table 3. pEP0cos6 region gene annotation summary (continued).

GENE	S'	3' # BP	#AA	MM (DA)	BEST MATCH [SCORE]	PROBABLE IDENTIFICATION
ORF1	1 ^a 731	732 ^b 244	26218	ACOA_BACSU [e-22]	acyl-CoA dehydrogenase	
ORF2	731 2275	1545 515	56317	AF047828 [e-94]	peptide synthetase	
ORF3	2301 3356	1058 352	39092	U57065 ZnaR [e-29]	antibiotic resistance gene	
ORF4	3412 4050	639 213	24819	(no match)	hypothetical protein	
ORF5	4248 12101	7854 2618	283282	PKSK_BACSU [e-190]	polyketide synthase	
ORF6	12097 14322	2226 742	80545	283857 ppsE [e-111]	polyketide synthase	
ORF7	14178 ^c 16451	2274 758	82007	(no match)	hypothetical protein	
ORF7.1 ^d	15866 14196	1671 557	61320	Y4HP_RHISN [e-28]	hypothetical protein	
ORF7.2 ^d	16507 16154	354 118	13102	Y4HO_RHISN [e-13]	hypothetical protein	
ORF7.3 ^d	16815 16507	309 103	11079	Y4RG_RHISN [e-13]	hypothetical protein	
ORF8	16830 17630	801 123	29633	PKSL_BACSU [e-2]	polyketide synthase	
ORF9	17628 25244	7617 2539	279286	pir1873015 [e-160]	polyketide synthase	
ORF10	25235 27877	2643 881	97101	AF047828 [e-111]	peptide synthetase	
ORF11	27867 32498	4632 1544	165084	AF091251 [e-167]	polyketide synthase	
ORF12	32498 39922	7425 2475	267116	AF040570 [e-223]	polyketide synthase	
ORF13	40031 45559	5529 1843	71258	AF091251 [e-119]	polyketide synthase	
ORF13.1	45599 46018	420 140	14590	(no match)	hypothetical protein	
ORF14	46015 47577	1563 521	55671	PKN1 MYXXA [e-34]	Serine/Threonine-Protein kinase	

^a The predicted ORF1 gene and gene product is truncated due to cloning of the DNA into the cosmid vector.

^b 731 is the last nucleotide of the last amino-acid-encoding codon; 732-735

5

36

10

15

20

25

30

35

40

45

50

55

is TGA (termination codon). Termination codons have been excluded in the present annotation.

c note: the predicted N-terminus of ORF7 shows 145 nucleotide overlap with ORF6

d note: ORFs on opposite strand - make up a transposable element from bp 16863-14130 (2733 bp, 11 bp terminal inverted repeat) that is similar to IS1131 from *Agrobacterium tumefaciens* (IS-66 like element, 2773 bp, 4 ORFs, 11 bp inverted repeat)

Table 4. p8POcos8 assembly analysis summary (continued).

a. p8POcos8 assemblies

ASSEMBLY	SIZE (bp)
Contig43	1017
Contig44	1246
Contig48	978
Contig49	1969
Contig50	2877
Contig51	2319
Contig52	1883
Contig53	4871
Contig54	7257
Contig55	5021
Contig56	10945
Contig57	16154

b. selected ORFs from Contig 56 and 57 of p8POcos8

gene	5'	3'	θ	bp	faa	best match (score)	probable identification
Contig56_003	3	8675	8673	2890	ERY2_SACER (e-300+)	polyketide synthase	
Contig56_027	10784	8682	2103	700	pir1160218 (e-300+)	transposon Tn1000 (E. coli)	
Contig57_001	92	1210	1119	372			
Contig57_002	222	7001	6780	2259	AL021899 plx12 (e-267)	polyketide synthase (M. tuberculosis)	

38

		CPKK_SACER [e-50]	Cytochrome P450 monooxygenase
Contig57_026	7424	6975	450 149
Contig57_027	7116	8375	1260 419
Contig57_030	9074	8337	738 245
Contig57_031	8412	9080	669 222
Contig57_032	8432	9667	1236 411
Contig57_033	9724	8600	1125 374
Contig57_037	10621	9755	867 288
Contig57_039	11085	10618	468 155
Contig57_040	10687	11091	405 134
Contig57_041	11887	10904	984 327
Contig57_043	11520	12998	1479 492
Contig57_044	13730	12018	1713 570
Contig57_047	13093	13797	705 234
Contig57_048	14064	13138	927 308
Contig57_049	14371	13535	837 278
Contig57_051	13900	14394	495 164
Contig57_052	14036	14440	405 134
Contig57_053	14678	15715	1038 345
Contig57_056	15211	15900	690 229

unknown ORF (S. lividans)

AF072709 [e-42]

Table: 5. AS assembly analysis summary (continued)
a. pEPOcos8 assemblies

contig	bp	ORF	5'	3'	bp	G+C content pos.3	best match [score]	probable function	identified domains using Pfam
10	9435	1	2861	582	2280	88.6 %	AAC44128 (e-170)	peptide synthetase [Sal/amylin]	AMP-binding aa 68-512; pp-binding aa 631-697
		2	3525	2802	824	87.0 %	no match	OmpA like	
		3	6420	4195	2226	85.1 %	P38370 (e-04)	unknown	
		4	7734	7135	600	98.0 %	P39760 (e-21)		
		5	9434	7950	>1495	96.2 %	no match		
11	18023	1	2526	751	1778	97.1 %	CA836518 (e-134)	peptide synthetase	AMP-binding aa 118-520;
		2	3588	2501	1086	94.5 %	CAB38516 (e-12)	peptide synthetase	DUF4 (domain of unknown function, found to the carboxy side of pp-binding sites) aa 5-360;
		3	5210	3579	1632	97.8 %	SS3993 (e-102)	peptide synthetase [Pyoverdine]	pp-binding aa 1-30; DUF4 aa 51-442; conflict ca. 20 aa at N-terminus are missing (conserved serine residue) conflict!
		4	8261	3811	2649	64.4 %	CAA11039 (7e-65)	polyketide synthase [Rifamycin]	ketosyl-synthase aa 14-439;
		5	8687	8258	1830	94.8 %	CAB05094 (e-144)	polyketide synthase [Phenolphthalein]	AMP-binding aa 371-758; pp-binding aa 840-904; DUF4 aa 937-1318; AMP-binding aa 1438-1825; pp-binding aa 1907-1071; DUF4 aa 1969-2388; AMP-binding aa 2485-2889; pp-binding aa 2872-3004; pp-binding aa 3027-3046;
		6	17320	6090	9231	92.8 %	AAD04757 (e-180)	polyketide synthase [Myosin]	DUF4 aa 43-476;
12	15898	1	1	1545	>1545	97.5 %	AAC68816 (4e-74)	peptide synthetase [FK506]	
		2	2883	1549	1335	94.4 %	P43492 (e-07)	cytochrome P450 enzyme	
		3	4659	2911	1749	92.5 %	NP 001130 (5e-56)	lipoygenase	
		4	5898	7066	1191	95.2 %	no match		
		5	7094	7822	729	90.1 %	no match		
		6	8974	7843	1032	87.8 %	no match		
		7	12001	11252	750	92.8 %	no match		
		8	13533	12181	1353	93.3 %	no match		
		9	15897	13981	>1917	93.9 %	CAA19149 (e-08)	regulator	

43388

Table 5.

b. selected ORFs from cosmid A5

gene	5'	3'	bp	kaa	best match [score]	probable identification
Contig10_001	2861	582	2280	759	U24657 saframycin (e-155)	polyketide synthase
Contig10_006	2880	3323	444	147		
Contig10_007	3876	2902	975	324		
Contig10_008	2953	3957	1005	334		
Contig10_009	4357	3299	1059	352		
Contig10_013	4180	6585	2406	801		
Contig10_016	7033	5600	1434	477		
Contig10_017	7043	6138	906	301		
Contig10_018	6500	7219	720	239		
Contig10_019	6954	7295	342	113		
Contig10_020	6982	7797	816	271		
Contig10_021	7819	7040	780	259		
Contig10_023	7415	8029	615	204		
Contig10_024	7794	9435	1642	548		
Contig10_027	9435	8806	630	209		
Contig11_001	770	417	354	117		
Contig11_002	2526	751	1776	591	AL035640 (e-113)	peptide synthetase
Contig11_004	1033	2787	1755	594		
Contig11_005	3500	1977	1524	507		
Contig11_007	3586	2501	1086	361		
Contig11_008	2507	3814	1308	435		
Contig11_011	5213	3579	1635	544	AF047828 syringomycin (e-86)	peptide synthetase

5

10

15

20

25

30

35

40

45

50

55

41

Contig11_012	6459	3811	2649	882	AJ223012 (e-52)	rifamycin	polyketide synthase
Contig11_016	6511	5210	1302	433			
Contig11_018	5551	6735	1185	394			
Contig11_021	9899	6258	3642	1213	283857 ppsE	M. tuberc (e-129)	polyketide synthase
Contig11_026	17329	8090	9240	3079	AF047828	syringomycin (e-300+)	peptide synthetase
Contig11_048	16733	17365	633	210			
Contig11_049	17397	17723	327	108			
Contig12_001	1	1545	1545	514	AF082100 (e-63)	(Streptomyces)	peptide synthetase
Contig12_002	1368	1	1368	456			
Contig12_003	3	1655	1653	550			
Contig12_005	2317	1361	957	318			
Contig12_006	2883	1549	1335	444			
Contig12_007	1777	4926	3150	1049			
Contig12_009	4659	2911	1749	582	LOX5_MOUSE (e-54)		ARACHIDONATE 5-LIPOXYGENASE
Contig12_011	4523	5065	543	180			
Contig12_012	4638	5231	594	197			
Contig12_013	4942	5520	579	192			
Contig12_014	6056	5541	516	171			
Contig12_015	5765	6373	609	202			
Contig12_016	5896	7086	1191	396			
Contig12_017	7095	5899	1197	398			
Contig12_018	5955	7331	1377	458			
Contig12_020	7549	7010	540	179			
Contig12_021	7094	7822	729	242			
Contig12_022	7995	7318	678	225			
Contig12_023	8408	7716	693	230			
Contig12_024	7916	9550	1635	544			

5

42

10

15

20

25

30

35

40

45

50

55

Contig12_029	10751	9732	1020	339
Contig12_030	9795	11132	1338	445
Contig12_033	10880	12124	1245	414
Contig12_035	12184	11252	933	310
Contig12_036	12079	14098	2020	674
Contig12_039	14098	12248	1851	616

Table 6. Sau4 assembly analysis summary.

a. plasmid Sau4 assemblies

<u>Assembly size (bp)</u>	
Contig17	2581
Contig18	4015

b. selected ORFs from cosmid A5

gene	5'	3'	# bp	lan	best match (score)	probable identification
Contig17_001	2485	1	2485	029	U24241 Sorangium (e-213)	polyketide synthase
Contig18_002	2	1510	1509	502	U24241 Sorangium (e-105)	polyketide synthase
Contig18_010	1494	4015	2522	841	U24241 Sorangium (e-245)	polyketide synthase

Table: 7. ORF data summary from A2 insert

ORF	5'	3'	bp	G+C content pos.3	best match [score]	probable function	identified domains using Pfam
1	1666	1	>1666	94.4 %	P54744 [5e-37]	regulation [serine/threonine protein kinase]	kinase aa 47-294;
2	1605	3338	1734	90.7 %	no match	acyl-IRNA synthetase	IRNA-synthase aa 27-694;
3	6100	3398	2703	96.0 %	CAA15124 [2e-180]	monooxygenase	oxidoreductase FAD/NAD-binding domain aa 110-227;
4	7111	6374	739	94.7 %	AAC32457 [2e-17]	aminotransferase	aminotran_1 aa 2-385;
5	9590	8433	1158	74.4 %	CAB42045 [9e-67]	L-dopa decarboxylase	pyridoxal deC aa 48-411
6	11393	9855	1539	85.5 %	AAD21754 [e-113]	oxidoreductase	short chain dehydrogenase domain aa 53- 240;
7	13656	12712	945	86 %	CAB41201 [7e-41]	polyketide synthase	ketosyl-synthase aa 13-438; acyl transferase aa 533-854; short chain dehydrogenase domain aa 1159-1357; pp binding aa 1451-1515;
8	15374	19584	4611	87.8 %	CAB06094 [e-180]	polypeptide synthetase	DUF4 aa 17-409; AMP-binding aa 507- 905; pp-binding aa 991-1054; DUF4 aa 1067-1466; DUF4 aa 1544-1844; AMP- binding aa 2041-2439; pp-binding aa 2525-2589;
9	20003	27889	7887	88.0 %	AAC80285 [e-180]	peptidase	Sigma70 ECF aa 17-83;
10	28251	29402	1020	86.6 %	BAA13079 [2e-44]	sigma factor	HTH aa 511-109;
11	31720	30401	1320	98.9 %	no match	regulation	response reg aa 46-159; signal aa 326- 542; response reg aa 590-703;
12	31982	32932	951	95.3 %	CAB09733 [2e-63]	regulation	
13	33128	33613	486	95.7 %	no match	regulation	
14	33661	34077	417	94.2 %	CAA19900 [e-37]	regulation	
15	35811	35255	357	80.8 %	CAA19859 [3e-15]	regulation	
16	37856	35730	>2127	88.8 %	BAA17685 [2e-50]	regulation	

Claims

5

10

15

20

25

30

35

40

45

50

55

5

45

Claims

10

1. DNA sequence, the expression products of which cause an enzymatic biosynthesis, a mutasynthesis or a partial synthesis of polyketide or heteropolyketide compounds or are involved therein.

15

20

10

2. DNA sequence according to claim 1, wherein the polyketide or heteropolyketide compounds are epothilones.

25

15

3. DNA sequence according to any of the preceding claims, wherein the DNA is derived from myxobacteria.

30

4. DNA sequence according to any of the preceding claims, wherein the DNA is derived from Sorangium strains.

35

20

5. DNA sequence according to any of the preceding claims, wherein the DNA is derived from Sorangium cellulosum.

6. DNA sequence according to any of the preceding claims, wherein the DNA is selected from the group consisting of:

40

25

(a) the following DNA sequence:
Seq ID No 1 (A2 cosmid)

45

30

GGATCGCGGCGCCCTCGCGCTGCTCCTCGAGCGTGCGGAGGAACTCCCACGCCAGGCCCGACT
TGCCGAGGCCAGGCGCGCCACCACCACCACCGCGTTCGCGGAGGGCTCGTCGACGCAATGGC
GCCACTCGGTGCGGAGCTGCGAGAGCTCGCGCTCCCGCCCCACGCAGGGCGTCGGCTTGCCGA
GCAGCCGTGGGACGGCATCCGGCTCCTCCTTCGGGCCGCGAAGCCAGCACCTCCGGGCCCT

50

55

5

46

10

15

20

25

30

35

40

45

50

55

GTACCGTCTCGAAGCGGCTCGCGAGCAGGCTGGCCGTCGCGTCGTCGAGCCGGATCTCCGGCG
GCGACAGGCCATCTCGCCCGCGATGAGCTGCGCGACCCGATCGACCAGCTCGCCGACCGGCA
GCCTCGCCTCGACCTCGGCCAGCCCTGTGCGGACGGACACGGGCACGCCTCCGAGCGCCGCCC
GCAGCGCGAGGGCGCAGTGGGCCGCCCGTGTGGCGAGATCCGTGGGCGACTCGGCGCCGGACA
5 GCGCGACGAGCCACCAGCGCGCTTGACGCCGATCGAGGCGCCCGCTGGCGCGCCCGATGT
15 CCCGACGCGCTCGGCCCGCGCGGCCGCTCGTCCTCCGAGAGCGTGGCGCCGGCCTCGGCGC
CGCCGTCTTCGGCCAGGATGACGCACATCACCTTGCGCTCGGCCGTCGTGATCGCCTCGCCCG
GCGCGGCCCGCGCCGCGACCGCGCTCGCCCGATCGAGAGCCCTCGCCGGCCACGGCGGCGA
GCTCCGCCCGGGCGGCGGCCGCTCGCGCGGCCGCTCTCCCGCTTCTTCGCCAGCATCCGCG
20 10 CCACCAGGCGCTCGAGCGGCTCCGGGATACCGTCGCGGAGCTCCCGAGCCGCGCGGCTCTT
CCAGGACGACCCGCATCAGGAGCGCGAGCGCGCTGTTGCCGAGGAACGGCGGGCGCCCCGCGA
GGCACTGGAACAGCACGCACCCGAGCGGAACACGTGCGCCCGGGCGTCGACCGGCGCGTCGC
25 CGCGCACCTGCTCGGGCGCTATGTACCCGGCGTGCCGAGCACGGCCCCGGGCGACGTGAGGG
TCGGCGGAGCCGAGGTGGCGCGCGATGCCGAAGTCGAGCAGCGTGACGCGCTCGACCGCGC
15 CGCCACGAGCATCAGGTTGCTCGGCTTGAGGTGCGGTTGAACGACGCCGAGCCAGTGGATCG
CGCCGAGCGTCGTGGCCACGCGCGGCCGAGGCCACGCTCTCGGCCAGCGTGAGCGGCGCCC
30 CGGCGAGCCGCTCCTCCAGGGTCACGCCGTCGAGCCACTCCATGGCCAGGTACGGCCGCCCTG
CGCCGGTCACCCCGTGCGCCACGTACTGCACCACGCCGGGAGCCGAGCGTCACGAGCGCCT
CCGCCTCCCGCGCGAACC GGCGCAGGTGCTTGGCGCTCGCGCCCTGCAAGACCTTGAGCGCGA
20 CCGCCTGCCCCGACACCCGGTCGCGCGCCCGGTACACGTCCCCCATCCCGCCGAGACGGCGA
35 GCGCCTCGATCTCGAAACGATCCTCGATCACATCCGCTGCGCGCATGGCGGTGCCAATGTACT
CCGCGCGAGCCTCGGGCCCCCGCGCGTAAGTGCGGCCCTGCGCCCGGTTGAACGCCAGCCCGA
GCGTGACCGCCTCGCGCTCGGGATCCACGGCCGCCGATCGGTCCACGCCTCGACGAGCGCCT
40 GCGTTGAACAACCCGCCACCGGGCGCACGCAGCCGGCATCGCCGCGCTGGCCACCCGGCGCTG
25 CCGCCCTTAGGCTCACCTCCGCGATGCCCGCTGGTTCAACACGGCAGGTCCCTGCAACCCGG
CCGATCACTACATGCTCCCGGCCGAGGAGCGCTTGCCCGAGTGC3CGATCTGGTCGATCGCA
AGGCCTACTTCGTCTGACGCCCCGCGGCAGATCGGCAAGACGACCTCGCTGCGCACGCTCG
45 CCCAGGATCTACGGCCGAAGGGCGCTACGTGGCCGCTCCTCGTCTCGGCGGAGGTGCGGCC
CCTTCTCTGACGATCCCGGCGCGGCCGAGCTCGCGATGCTCGCAGAATGGCGCGGCACCGCCG
30 GCGCGCAGCTCCCGCCGATCTGCGGCCGCCACCGTTCCCGAT3CGCCCGCCGGTCAGCGCA

5

47

10

15

20

25

30

35

40

45

50

55

TCGGGGCCGCCCTGCGCGCCTGGGCTCAGGCCGCGCGCGCCCGCTCGTCGTCTTCCTCGACG
AGGCCGACGCCCTGCGCGACGCGACGCTCGTCTCCCTATTGCGCCAGATCCGCAGCGGCTATC
CCGACCGCCCGGTGACTTCCCGCACGCGCTCGCCCTCGTCGGCCTGCGCGACGTGCGCGACT
ACAAGGTCGCGTCGGTCGACAGCGGCAGGCTCGGGACGTGAGCCCCCTTCAACATCAAGGTCG
5 AGTCGCTCAGCTGCGCAACTTCAACCGCGACGAGGTGCAAACTCTACGCTCAGCACACGG
15 CCGAGACCGGTCAGGTCTTCCGGCCGACGCCGTGGACCGCGCCTTCGAGCTCAACAGGGCC
AGCCGTGGCTCGCCAACGCGCTCGCCCGCCAGCTCGTCGAGGTCTCTCGTCAAGGACCGCGCGC
AACCCATCAGCTCTGCGAACGTGATCGCGCCAAGGAAATCCTCATCGAGCGGCAGGACACAC
ACCTCGACAGCCTGGTGGATCGGCTGCGCGAGCCGCGCATCCGCGCGGTGATCGAGCCGATGC
20 TCGCCGGCACCGCGTTGCCGAGCCTGCCCCCGACGACCTTCGTTTCGCGATCGACCTCGGCC
TCGTGCGCATGACCGCGGAGGGCGCCCTCGACGTGCGCAACCCCATCTACCGCGAGATCATCG
TCCGCGAGCTCGCGTTCCCGATCCCGCCTCACTCCCCCAGATCAAGGCCACGTGGCTCACGC
25 AGGACGGCCGCTCGACCGGACCGCCTGCTCGACGCTTCTCTCTCTGCGGCCAGCACG
GCGAGCCGCTCCTCGCGCCGCGCCCTACCATGAGATCGCCCCGCACCTCGTGGTGTGGCCT
15 TCCTCCACCGCGTGGTGAACGGCGGTGGCACCGTCGAGCGCGAGTACGCCATCGGCCGGGCA
GGATGGATCTCTGCGTTCGTTACGCGGGCGAGACGCTCGCGATCGAGCTCAAGGTCTGGCGAG
30 ACGGCCGCCCCGATCCCGTCGCGGAGGGGCTCGCCCAGCTCGACGAGTACCTGGCCGGCCTGG
GCCTCGATCGCGGATGGCTCATCTCTTCGACGAGCGCTCCGGACAGCCCCCATCGCCGAGC
GCACGCGCCGCGAGCGCGCTCTCCCCCGCGCGCGAGGTGCGCGTCATTGCGCGCTGAG
20 GGAGCTCGCCGCGCGGCGAGCGCCCTCCACGAGGGCCGGCCACCTCGGACAGCGTCTCTACT
35 CCTCCGAGGCCGCGCGCCCCCGCGCCGCGCGCGCGCGCGCGCTCCAGCTCGCAGC
GCACCACCAGGACCTCGCCATCCGCGAGCTCCGGCCGCTCCACGAGCGCGTGCAGCGCCGCGC
GCACCGCGGTGAGCACGTCTCCAGCGCGCGCTTCAGCCGCGCCAGCGTCGCGGCGTTCCGCC
40 CGAGCGCGAGGTGCGTACGACGCGCCCCACGCTCGCGCCGAGCTCGCTCTTGCGCTTGTTGA
25 CCGCCGCCATCGCCCGCGCCGAGATCCAGGAGCCCCGATCCGAAGGCGCGCGACCGCCG
CGAAATCCGCCGTGAAGGCCACTTCGCCCCGTGGATCGAGGTATCGCCCCGTCTCTCCGCGT
ACACCCAGCGCCAGACCTCGTCGGTGATGTACGGCAGGACCGGCGCGAACAGCCGACGAGCA
45 CCGACAGCCCCGAGCCGAGCGCCGACCGCCGAGCCGCGCGCGCTCCCCGGCGCGCCCT
CGCCGCGCGCCCGCGCTTCGCGAGCTCCAGGTAGGCGTCGGTGAACGAGCGCCAGAAGAAAT
30 CCTCGGTCCGCTCGAGCGCCGCGCGAACTCGTGCTCGTCGAACGAGCGCGTCCGCTCGTCCA

5

48

10

15

20

25

30

35

40

45

50

55

CCACGGCCGACAGCTTGTGCAAGAGCGCCCGGTGAGCTCCTCGGAGATCGGGTGGACCTCCG
CCGACTGGCTGAGCACGTACTTGCTCGCGTTCCAGATCTTCGTGACGAGCCGCTTGCCGATCT
TCAGCACCTTCTCGTCGAACGCCGTGTCCGTGCCGAGCCGCGCTCGCCGACCAGTAGCGGA
CCGCGTCCGAAGAATACGTGTCGAGCAGGTGCATCGGCGTGACGACGTTGCCCTTGCTCTTCG
5 ACATCTTCTTGCGATCCGGATCGAGGATCCACCCGAGATCGCGACGTGGTGCCACGGGACCG
ACGACTCGTGACGATCGCCTTCGCGATCGTGTAGAACGCCACGTCCTGATGATGTCTGTTGG
15 CCTGCGGCCGAGATCGGCCGGGAAGAGCCGCGCTGGCGCGCCGGATCGTCCCCCAGTGAG
AGCTGATCTGCGGCGTGAGCGAGCTCGTGAACCACGTGTGAAGACGTGCGACTCGGCGGTGA
AGCCGCCGGGCTGGTCCCGCTGCGACGCTCGTACCCGGGCGGCACGTGACCGTGGGTGCGA
20 CCGGGAGCATCTCGCGCTCGCGAGCAGCGCCGGCTGTGATCCGGGTGCCCTCGGCGTGA
GCGGATACCAGACCGGGAAGTGCACGCCGAAATACCGCTGGCGGCTGATGCACAGTCAACCT
GGAGCCCCCTCGGTCCAGTTGCGGTACCGGAGGCGCATGAAATCCGGGTGCCACTTGATCTTGT
CGCCGTATTTCGAGGAGCTCGGCCTTCTTGTCGCGAGCCGGACGAACCACTGCCGCTGGGCA
25 CGAAGTTCGAGCGGCTGGTGCCTCCGCTCGTAGAACTTCAACGCGCGCTCGATCGGCCTCGGCT
15 CGCCCCGAGCGCGGCCCGCCGGCGCGCGCGCTGCTCCTCGCGGCGGAGCAGCTCGA
CCACCGCCGCGCGCGCTGCTTACCCCCCTGCCCTGGAGCGGCGCATACGCGGCGTTGGCCG
CGGCCGGGTGCGGCTCTCCACGCGCCCTCGCCGAACGTACCGGCAGGACACGGCCGTTCT
30 TGCCGAGCATCTGCCGAGCGGGAGCTTCTGCTCCCGCCACCAGATCACGTGGTTCGCTGCG
CGAAGGTACAGACCATCAGGATGCCCCTGCTTCTCGCGATCCACGAGCGGGCTCGGGAAGA
20 TCGGCACCGGCGCGCGGAAGATCGGGGTGAGCGCGCTTTCGCGAAGAGGTGCTGATACCGG
35 GGTCTCCGGGTGCGCGGTGACGCGGACGCGAGCCGCGAGCAGCTCCGGGCGCGTCTGGCGA
TGACGAGCTCCTCGGCCGTCCCTCCACCGGAACGCGATGTCTGGAACGCGCCGATTGCG
GGCGATCCTCGACCTCCGCCTGGGCGACCGCGGTCTGGAAATCGACGTCCACATCGTGGCG
CGAAGACCGAGTAGAGGTGGCCCTTCTCGTGGAGATCCAGGAACGACAGCTGCGCCGCTCTGC
40 GGCAATGATCATCGATGGTGGCGTACTCGTTCCGCCAGTCGACCGAGAGGCCACCCGGCGGA
25 AGAGCGCCTTGAAGACCTGCTCGTCTCGCGCGTGACCTTGTGGCAGAGCTCGATGAAGTTGG
GCCGCGACAGATGCGCGGCGGCTCCTTCTTGATCGTCTCCGGCGCGGCTGCGGCAAGGTCA
45 GGCCGCGCTCGTACGGCGTGCGCACGTGCGTGCGGACGTGGAAGTAGTTCTGCACGCGCGCT
CGGTGGGCGAGCCGTTGTCTCCAGCCCATCGGGTAGAAGATGTTGAAGCCGCGCATCCGGC
30 GCTGCGGACGACGACGTCCTGCTGCGTGTAGCTGAAGACGTGGCCGATGTGCAGCGAGCCCC

5

49

10

15

20

25

30

35

40

45

50

55

AGGCGGTCGGCGGCGGGGTGTCGACGACGAAGGTCTCCTCGCGGGGCGCGACGGGTCGTATC
GGTACGTCCCGTCGGCCTCCACAGGTCCGCCAGGCGCAGCTCGGCCACGGGCGAGTCGAAGT
GCTTCGGGAGCGTCGCGGGATCGATGGAGCGGAACGTCTTCTTGATCGTCACGTGGTCACCTG
CAGAACAGACCCCGCAGGAACCGCCCGGGGCGCGCATCCTACGTCTGTCCTCCCGGGTGCCCGC
5 TCAAGGCGCGCCGCGCCGCGCGCGCGCGCGCTCCGTGGAGCCGAGAGGAGAGGCCGCGCCAGGTC
GGCGGCTCCGCGGCGCGCGCGCGCGCTCCGTGGAGCCGAGAGGAGAGGCCGCGCCAGGTC
GTGGAGGACGCGCGGCGCGCGCGCGCGGAGATCGCGGAGAGGCGGGCGCATCGATCGCGGCCA
GGCCGGGGGCTCAGTCGTAGCGCTCGACGTGGACGTGCTTGCGGTGGACGCCGAGCTCGCCGC
GGCGGAGCTCGCGGACGGACGAGACCATCCGATCCAGGCCGAGATGAAGACGTGCGGCGCCG
10 GATCTCCGCTCTTCTCCGCGAGCTCCCGGTAGAGCTCGGGCACGTGCGCCTGCACGTAGCCGC
GGCGGCCGGCCACGACGGGCCGCGCGCGAGAGCGTGATCTCGTAGCGGATCCGGTCGGATC
CGCGCGGAGCGCCTCGAGCTCGTCGCGGTAGATGACGTCTCCTCGAAGCGCGCGCGAACA
GGATCCACAGGTGGGGCGCGGCCAGCCCGCGCGCAGGAGGCGCGCAGCATGCTCCGAGCGC
GCGTGATGCCGGTGCCGGTCGCGACGAACAAGGAGGGCGCGGAATCCCCGGGATCGCGGGTGA
15 AGAGCCCGTGCGGGCCGATGGCGCGGAGCGTGCGCCGGGCTCGAGCCGGTGAGGTGCTCCG
AGCCCCCCCCCGCTGCACGAGCGTGACCGCGAGATCGAAGCGGGGCGAGCCGTGCGGCGCGG
ATGCGATGGAGTAGGCGCGCTTACCTCGCCGCCCGGGAGCGGGAGGACGAGGTTGACCCACT
GGCCCCCTCGAACAGAAACGACCTCCCGTCGGCGCGCTCGAACGAGAGCTCGCGCACGAAAG
GGCTGAGGGGCGGGCGGCGACGAGGCGGGCTTCGAACGGTTGCGCGTGATCATGGTCGGGG
20 CCCGGCGGGGCTCGGCTGCGAGGCCGCGCGGGTGGCGAGGTCTTACCGCAGCCTGCGCCCCG
CCCAATCGCGATCGCCGCGGAAGGGCGCCGCGGAGGGCGCGCAATCGCGGGAATCACGGGC
TTCCGCCCCGTGCGCCGCGGAGCGCGCGGGCGGGCCCGGCCCGCGCTCCGGCGGGGAGCC
GTCGCGGGCTCTACCGCACGCCCATGCGGCGGCGCTGCGGGATGTTACCGCCGGCGGGAGC
GATCCTGGTTGGGGAGCGCGCGCGGGCGGGGATCCCGGTGCGCGGGCTTCTGCGCGGGGA
25 GCTGCCCTCGCTGAGCCGGGCGCTGGTCGGGCGACTTGCCGAGCCCAGCGCGAGATCGGAGA
CGGGGAGATGCGCGCGTCGCTGCATAGAATCCTCCATGGAATCGGTCATCAACACATCGGGAA
GAGCACCCAGGTGAAAGAAACCTTCGAAGAACCGGCTCTCATACCCCTCCATTTCATCGTGC
45 GACCCCGGATTGAGGACGGATCGAACCCGCGAGGGACGCTGGCTCTCTGGGCCTCTCCCTGCT
CGCTCGACCGGCGCCCTCTCGACGCAACTCCGCCGTTCTGTCGGGACGGGACGGTCCGCCTCGC
30 CGCACGCTCCCGTCGAGACGACTCAGCCTCTCGACGTCAGGAGAGATGACGACTCGGCCCGT

5

50

10

15

20

25

30

35

40

45

50

55

CGCGCCACGACCCCTTCCGGCTCGGTGCTTCGAGCGCGCGGCCAGCGAGCGAGGGGCGATCGCC
AGGAGATCACGAATCTCCCGGCCATCGGCCTCCAGCGCCTCGGGCTCGTTTCGCTCGTCGCCCC
GCTCCGTCCCCGCGCGCGCACGACGCGAGCTCGCGCGGGGAACCGCGGGCCGCTGTCTGTGGCT
GCTGATGCGCGACGATACAGGGGGACGCCGTGCCTACCTGGGCAACAGGCGCTCATCTTCTA
5 CCACGGCGAGCACTACGGTGAGTGCTGCCATGAGTAGGCCCTGAGGGTCCGCGCGACGGAGC
15 GTGGTGTCAGCGAGAGATGCGCATGGTGACGCGGGCTACGCGTCGAGAGGGACACTAGCACT
CGACCTCGATCCTGCCCAGCACTTTTTGTTCGGGGAGGGCTGCCCTCCCGCTGGCCGCTGGCCG
CTGGCCGCTCGCCGCTGGCCGCTCGCCGCTGGCCGCTCGCCGCTGGCCGCTGGCCGCTGGCCA
TGTCGACGTCGAGCTCGAGCAGCCCCGGGCTGACGGACAGACCCCGAGTTTCATCGAGCCGGT
20 GATGCCGAACCCGCCAAGCGAAAAACGTATCCGTTTCGGCAGGTCGTGGCCTATCATGCAAGC
TGCTCGATGCGCTGACAGGCTTCTTCGAGATCCTCGTCGGTCTTTGCGAAGCAAAACCGCATG
AAGCGACTCCCCTGCGTCCCTTCAAAGAAGGCGTCGCCTGGCACGCCCCGCCACCCCGGTCTCG
25 TCGAGCAAGTAGATGGCTCGCTCTCGACCTGTCTCCCGGGTAGGCGAGACACATCCGCCAGC
ACGTAGTACGTCCCCTGCGGCACGCAGGGTGGCAAGCCCGCTTTCTCCAGCGCCCGACAGAAC
15 CGGTCTCGCTTCCGTTTCATATCCCTGGGCAAGCCCCGTGTAACAGAGCGAGGAAGGCCCGG
ATCCCGGCGAGCGACTCCATGCTGCAGCGGCGTCGGCGCGCAGACATACAGCAGGTCGCTCATG
30 GCTCCAATGGCCTTCGCCCACCTGGCATCGGCCACGCTGTAGCCGATCCGCCATCCTGTGATG
CTGAAGGTCTTCGAGTAGCCGCCTATTCGTGATCGTACGCTCGGACATGCGCGGAAGGGAGGCG
ACGCTGACGTGCTCACGGCCGTCGAAGATAAAGTACTCGTAAATTTTCGTCCGTGATCACCATG
20 AGGTTCATGGTGGCAGGCGAGATCGGCGATCTGTTCCAGCTCCATTTCGGCCGAACACCTTCCCG
35 GAAGGATTTCCAGGAGAGTTTACCACGATCGCCTTGGTCTTCGGGGTGATCGCGCGCTCCAGC
TCGTGCGCGTCGACATTCAGCTCAGGGATCGCGCCGTCACATACCGCGGAACAGCCTCGACG
GCGAGGATAGCCTGGGCGTGATAGGCATAAAACGGCTCGAAGAGCAGCACTTCGTCCCCAGGA
40 TTGAGCAAGGCCATGCAAGTGGCCTGAAAGGCCCTGTGCTCCGGCGCTCACCGTGATGTCA
25 GTCTCCGGATCCGCCCGGATGCCATTATGGCGAGCCAGCTTCGCCCGGATCGCATGGCGCAGC
TCCACGATGCCGTCGAAGCGCGAATATGTATTGCACCCCCGATCCATCGCCTCCTTCACCGCT
TGAAGGATCACCGAAGGAAGTGGGTATCACAGACGCCCTGGGACATATTGATCCCATGGACC
45 TTGGCGCACGCCAGGGTCATGGTACCGATATCGGACTGGGCGAGGCGAGCCGCACGATCACTC
GGTAGACTCTTCATCAGCGTGCTCCTGCTTCTGTTCTGCGGCTCTGCATGGTGTCTTCGGGTG
30 GGCTTGTCAGCTCGACGCGCCCATGCAGCGGCGCAGCCCTAGCGGCCGAGGTCTGTCCACAC

5

51

10

15

20

25

30

35

40

45

50

55

TTCTTTGATGAAAGCGAGACATTTCGCTTTTCGTGCCCTGTTTGCCCGCAGCCCTCCAGCCCC
AGGTACGGGCTTGTTCGGCGGGCCAGATCGAGTACTGCTCTTCGCCGTTTACCACGACCTGGCA
ACGCGTCTTGCTTTTCGTTCGTCCCGATTTCATGATTTTCCTCGCCCTTCGTTCAGCGCTGCGCGAG
CATGAAACGAATCGCTCATCGGCGCACAGGCGCGCGCCGGCTGCCCGAGGCACTCCCACGCC
5 TCCCTCACGGCAACCTCATCGCTCCGGATGTTCCCGATGGCGACTCGGATCGTGACCTGCCG
15 TGGAGACGGGTATGGGACAAAAATACCTGCCCGACTTGTGACCTCGTCCAGCAGCGCCTCG
TTGAGGCGATCGAGCTCGCGTTTCGATCGACTCTCTCTGCTCGTCCGCCGACCGCATGATG
CAAGCGAGCGCGGAGGGCTTCATGCGAAAGCAGACCGTACTGAACGGCGTCGGCGCGAGGCGC
TCCCAATCGGGATCGGCGTCCACCCACTGGGCGAGCTGCTGCCCAATCGGAGGTGCTCCCGG
20 ATCCGGGCCGCCAGCCCTTCATGCGCGAAGTAGCGCACGATCATCCAGAGCTTCAGCGCTCGG
AAGCGCCGACCGAGCTGGATACCCAGTCCATGTAATTCGTGACGTGCGCCCTCGGTGCGGAGG
TATTCGGGCACCAGACTGAACGCGCGCTTCAGTCGGTCGGCGTCACGCACGTAGAGCACGCTG
CAATCCATGGGGGTGAACAGCCACTTGTGAGGGTTCACTACCAGCGAGTCCGCCCCCTCGCAG
25 CCGCGAGCACGTCCCTGTGCTCGGGACGATCGCGGCCATCCCCGCGTAGGCCGCGTCCACG
15 TGAAGCCATAGCCCGTGCTCCCGGCAAACGCTGACGATGGCGGGGATGGGGTCGACGCTCGTC
GTGGACGTCGTGCCACCGTCGCCGCGACGCAGAGAGGGTCGGAGGCCGGCCCCGAGGTCTCTCC
30 ACGACGGCGGGCGCGCAGCGCCTCGGGGACCATGCGGAAGGCCGGATCCGTGGGGATCTTCCGC
ACCCCTCTGCCCCGATGCCGAGGGTGATGGCTGCCTTCTCGATGGATGAGTGCCTGCTCC
GACGCGTAGAGTCGCATGCGCCGCTGTCCCGCCATGCCCCGAGCCGGATGGTCGGCTCGGCC
20 GAGTCGCGCGCGGCCCGGATCGCGA.CCATGCTGGCGGTTCGACGCGGTGTCCATGATCGCGCCG
35 TGCAAGCCGGCGTCGAGATCCAGCATCTGACGCAGCCAGGAGGACGAGCTCTCGAGCTCG
GTGGCCCGCGGCGACGTGCGCCATAGCATCACGTTGACGTTGAGGCACGCCGCGAGCAGCTCG
CCGAGGATCCCAGGACCAGACGCCGTGTTTCGCGAAATACGCGAAGAATCGCGGATGATTCCAG
40 TGCCTGATCCCCGGCAGAATGATCTGCTCGAAATCGGTGAGCACGGCGTCCATCGGCTCCGGC
25 TCGACGGGCGGGTTCGGGGCCAGCCTGCCCTTCACGTCGCCGGGGCGGATCGCGGGAAGACG
GGGTATCGATCCGGGTGGCCGAGGTAATCGGCCGCCCAATCGATGATTCTCATACCGATCCGG
CGGAACCTCTCCAGATCCATGTCCCGAGCCGTTCTTTCCGCGGGTCGCTCACGTCAACTCC
45 TCGCCCTGCCAGGACAGGATCCTCGAGGTCCCCTGGCTCCGGCGGTGGAAAGCGCTCCTTGAA
CGTGAAGGCCACGGGGTCGGTCCCTAGCGCCGAGGTGCTCGAGCCGATCCTGCCCTCGCG
30 GACCGACGGGATGTGCCCGGCCGGGACCCACCACAGCACGAGGTAATGCGGCTCGAGATGCTC

5

52

10

15

20

25

30

35

40

45

50

55

GAACCACCGAGCGCGCTGTCGCAGGAACGCGGCATGATCCGCGGTGTAGGTGAAGGCGAACAG
GTGCTCGATGGAGGTCCATACCGACAGGGTCACGAGGAGCCGCTGGTCCGGGTACGGACGGAT
GGACACAGAGTTCCTTCGGCCGTCTGCAGGCGCCACACGAACCCCTCGCTCCGATCGGCCAG
ATGGTTGATATGGTTCGAGCCCTGGACGAAGCCCTCCATGATCGGATCCTCCAGCGGAGCGCG
5 AATACATGCGAAGTTGTATTGCGCGATGTGGTGCCGATGCTCCGACATGTCGCTTTCCATCTC
CAGCTCCCGCTCACCAATCCCAGCGCTGCTCCGGGGAGCTCATCAGGGCAGACGCGACATCGA
TCCCGAAGCTCCGCCGCATCCCTTCGACGAAGGCGGCCTGGACCGCTTCGGCGACGGATCGGC
CTGCCTCCGGCAAGACCTCGGAGACAAAGAAGAACCGCCTCGTGGAAGGGACAATCTTGCCCC
GCTCCGCCCTGGCGCCATACGAAGTGCCCTCGTCACCAGTCCCTCCGCGTCGGCATAACCGACCT
10 CGCCGGCGCCGACCGCCACGCTGCCGCCTGAGCCGAGCTCCACGAACGCCTCACCGCCTCCCG
AGATCTCGAGGCGAACGTCCGGGCCAGCCAGATCGCCGAGGTCCCAAGCGCCGACGGGGACGG
CGAACCGCAGCGACAGGAGGTTGTAAAAATCGACGAATGCGTTGATGTGCGGCAGCTCTCCAC
CACCAGGAGACCGCTTCGCCAGCGCCTCGATCGAGCTCGGAAATTTCTTGCCAGAGACCCCCA
CTCGCTTCATCGCCTCGCGCCAGGCAGCCACGTGCGGATGCGACTGGGCGTTTTCTGCGGCCCC
15 AGCTCCGTCGCAGCTCCTCCTCGACCTTCCGGAGCTCCTCCAGCACGGCCGGCCGCTCTGCGG
CGTTGTCCAGGCCTTCCCCGTACCCGGTGACCAAGATCATCCCAGGAAACGACTCCCAGATTCT
GCGGATCGACGATGAATGCCATGTGCCTCCTGCCCCCTCGAGAGCGATCGCCTCGATCGACACC
AGGCTGTGGATGCATGAGCCGGGCCGTGCGGACGCGAGACCCCGCTACTCATGGCTCTTCGTG
GCCGATGAACAGGTCTCCACCCGTCGATCGTGCTCGGTGCCCCGATCCGTCCAGTCCCACCC
20 GCCGGCGACCGCGATGTTTGACCCGAGACGTACGAGGCGCGGTTCGAGGGCGAGGAACGCCAC
AGCATCTGCGACCTCGCTGGCGCGCCCCAGGCGGCCCATGGGGACGCGCCGCTCCATCCACTC
CTTCTGCGCGGGCGGAAGGTATCCGTTGTGCGATGAGCCCTGGAGACACACAGTTGACCAGGAT
TCCATGAGGCGCCTCCTCCGTGGCCAGGCTGCGCGTGAGGATGAGCACGCCGGTCTTCGCGAT
CGAGTACGCCGCCACGTTTCGGCGCGCCGCGGATCGCGTACGTGGGGCTCAACCCGATATTGAT
25 GATCCGGCCGCTCTTTGCTGGCGCATGCGGCCACGGCCGCGCGACAGAGGTAATGAACGCT
GCTCAGGTTGCTGTCATGACGTTGCGCCATTGCTCGTCGGTTCATCGCCGCAAGCGGCTTGAA
GAAGAAGTCGCCCACGTTATTGACGAGGATGTCGATGGGGCCCAGCTGCGCCTCGACGCTGGA
GAAGAGCTCCGCGGCCGCGTTGGGGCGGGTGACGTCGGCCTGCACCACCATGGTTCGTCGCC
GAGCGCGCGGATCTCGGCCGCCGTCTGCTCGGCCGATCCTTGTTTCAATGGTAATTGACGGC
30 GACGTCCGCGCCTTGCTCCGCGAGGCGCAGCGGATCGCCTTGCCAATTCCGCGCGAGCTACC

5

53

10

15

20

25

30

35

40

45

50

55

GGTGACCAGGGCGACGCGCCCGGCGAGCTCCAGCGATCGCGCCTGTGGCAGGGCCGGAGCAGC
CTCCTGGTGGAGCTCGACGTCGACGGGGAGCTCCACGTGGTAGCTCGTCTCTCGCGGAGCCGC
GCAGTACCTCTCGTAGAACGCCTCGAGGACGGGCTCTTCGCGTCGCATGATGTCCGCGTGGGA
TTGGGCGCTGCGCCACGGATAGAGCACCAGGATCTCGTCGGGGCGCACGGTGCTCTGAAAGAA
5 TCGCGCATGACCGCACCACCCGGGTGTTCGTGGGCTCTGGGCGGAGCAGATCATCCATTTT
TTGCATGATCTGCGTGGCCTCGCCCTCCATGCCGGGTTTGATGCGCCATCGCTCCATAACGAG
15 GATCATGTCTTGGCTCCTGTTCGTTCATCGCCGTTTCGATCTGGGGGGGCTGCCCGCGCTCTCG
AGGGCGCGCCCTTGTATTGGCCCGGATGGTCTGGGTAGCGCTCGCGAGCTTTCGCTTGTGG
GCGGCGTTCAGGCTTGGCCTTGATTGACGAACCGCTCGCAGACGAATGCGTGCGATTTCATAT
20 GCGGTGCGGAGCGCCACAGGTAGAGACGGTGGCTACGAGCGAGCTTGGGCGGAAGCGCCCTC
AAGGCGGGTCGGATGGATTGAGGGTGCGAATCAAGCGGGCGTGCTCTTGAAGAAGAGACCG
CTGAATCCATCCTGGACGAACGGCGGGGCCATGGTAGGTGCGACGATGTTGTTGTAATCATCA
25 TTCGAGAAATCGCCGGTAAACCGGAAAGATGCCGCCGTTGCCACCAGAGCTGCGTGGAGATC
GCGAGCGCCTCTCGCGCCCTCGGAACGTACGGTTGGCGAGCGCCTCCGACAGCCACCTGTGC
15 GCCATGAGCAGCGCCTGGACCAGCACGAAGAACGCGTGATGCCCAATACCCAGCGGCCAGC
GCGCGCTCCGGAGCGCCCGCTCCGCGGAGGCGCAGCCAGCTGCGGGATGCCATTCCATGAT
30 TTTGGCCTTCGCTTGGCGGAGAACTGGTGGAGGATGTCTCGATGGACGCACAGAGATGCCCC
ATTTCCATGGGTTGACAGGAAGTACCTTTCAGGCTTTCGCGGATCATTGCGTAATATGCGACG
ATCACCGCTTCGCGAGTACGTCTCGAGCGGGTCTGGGGCGGAGACCCGGTGACATGGAAATAG
20 GCGTCGTAATCCGCTCCCGGTCCGACAGGCTCCCGGGATGTCCGGATCTGCCGCTTGCGCC
35 TGCCATCGCTCGAGGATGGGGACGGCGAGGGGGCGAGATGCTCGGCGAGCACCAGGTCG
CCCTCCGGCGCATGCGCCAACAAGACCTCGAAGGCCTCGGCTGCGACCTTGGAGGTCTCGCCCC
ATTCCGACGGCCTCGATGGCTTGGGGCAGCGGTAGATGGATGGTATATTTAGCCATGATTTGC
40 CCGAAGATTGCCGCTGCGTCGACAGATCTTTCGCGAGCCGGAACGCCATTTCCACTGCTCTGG
25 CTCTCAATATTGAATTGAGCCCTGGCGACTGCCATAGGCCAGTCGCTCGACACAGTGATACGG
AGCGGCCCCGATGCTTTCTCCTTTTTCAGTCTGACCGAATACTTCTGTTGGGCGCCAAAGAT
CCCTTGCCGAGACTGTCCGGCGAGATGTCTGTGCGAAGCGTCCGACGTCCAGCGGGCCCAT
45 GCGTTGCTAGAGCATAAAACGGTTCGATGCCTGGTTCGAGAGGGAGACGCGAGGAGCCTCCCTT
TGGGACGGATGAGGAATTTCTGTGACCGAAATGTCCGCAGGAACAGCGGCGCAGAAGCGGCGCA
30 TCGATGGGGAACCATGGGTACGAAGACATTGATGATAATGTGACGCAATCGCAATCGTCGC

5

54

10

15

20

25

30

35

40

45

50

55

GATGAGCGGCCGCTTCCCCGGCCGAGAAACGTCGAGGAGCTGTGGCAGAAGCTCCGCGCTGG
CGTGGAATGCGTCGTACCTTCACAAGAGCCGAGGCGCTCGCCGCGGGGTGAGCCGCGAGAT
GCTCGCGAATCCCAGCTACGTGCGCAGAGGCGCGCCGCTCGACGGCGTGGAGCTCTTCGACGC
CTCGTTCTTCGGGTTAGCCCCAGGAGAGGCGAGAGCATGGATCCGCAGCAGCGCATCTTCCT
5 GGAGGTCGCTGGGAGGCCCTCGAGCGCGCGGTTACGACCCCGATGCCCATTCGGGGCTAT
CGGCGTCTTCGCGGGCAGCGCCCCGAGCGGCTACCACTCCCTGGCGCAGTCCGACCCGGAGAT
CCTAGGCGCCCTCGGCCACTACCAACTGACGCTGAACAACGACAAGGATTATCTCACCACACA
CGCCTCGTACAAGCTCAATCTGCGGGGCGCCGAGCGTGTGCGTGCAGACGTCCTGCTCGACCTC
GCTCGTGGCGTGGTCAATGGCCTGCAGAGCCTGCTCAACCACGAGTGCACATGGCGCTCGC
10 GGGTGGCGTGGGATCCATGCGCATCAGCGGAGGGGCTATCTGTATCAGGAGAACGGCATCTC
TTCGCCCGATGGGCATTGCCGCGCTTCGATGTGGCCGCCAAGGGCACCGTGGGCGGCAGTGG
CATAGGCATCGTCGTCTGAAGCGGCTCGCCGACGCGCTCGCCGACGGCGACACGTCACGC
GGTGATTTCGAGGAGCGCGATCAACAACGACGCTCGAGCAAGATCGGTTACACCGCGCCGAG
CGTGAGGGGCGAGCCGAGGTGATCGCATGGCCGAGGCGCTCGCCGGCGTGGAGCCGATGA
15 CATCAGCTACATCGAGGCGCACGGCACGGGGACGCGGCTCGGCGATCCCATCGAGATCGCAGC
CCTCACGCGCGTGTTCGGGGCAAGACCGCACGAAGGCAGTTCTGCGCCATCGGCTCGCTCAA
GACCAACCTCGGCCACCTCGATGCCGCGGGCGTGCCTCGCTGATCAAAACGGTCATGGC
CCTCGAGCACCGCGAGCTGCCCCGAGCCTGCACTTCGAGCGTCCGAATCCGAAGCTCGAGCT
GGAGAGCAGCCCTTCTACGTCAACACCCGCTCACTCCGTGGCACGCGGCACGAGGTCCGCG
20 CCGCGCTGGCGTCAGCTCGTTCCGCACTCGGCGGCACCAACGCGCACGTGGTCTCGAAGAAGC
TCCGGCCCCGCTCCGAGCGGCCCTCGCGGCGTTGGCAGCTCCTACCCCTCGCGGCTCGCTC
CGAGCCCGGGCTCGCGGGGCCACGSCCGACATGATCGAGCACCTCGATCGCCACTCCGGCAC
ATCGATCGCCGATGTCACGTACACGACCGACGTCGGGGCGCGGGCTGGCCCTTCCGGCGAGC
GGTCGTGCGGAGAGCGCGCGGATCTCCGCGCGCGCTCGCGAGCGAGGGCTCGCCGCGCTC
25 GATCTCGTCATGCCAGGCGGCGAGGAGAGGCCCGTCGTCTTCTGTTCCCGGTCAGGGAGC
GCAGCACCTCTTCATGGCGCGGAGCTGTACGAGGTCGAGCCGATCTTCCGGCAGTCCCTCGA
CCGCTGCGCCGAGCTCCTGCGCGGCCGCTCGGCTCGATCTGCGGCAGGTCTCTACCCCGC
CGAGGGGCGAGCGGACGACGCCGAGCAGGAGCTCGGTAGGACCGCGATCGCCAGCCCGCGCT
GTTCGCCATCGAGCTCTCGCTCGCCAAGCTGTGGATGGCCTGGGGGATCGTCCCCAGGCGAT
30 GATCGGCCACAGCGTCGGCGAGTTCCCGCGGCTTGTCTGGCGGGCATCTTCCGCGAAGAGGA

5

55

10

15

20

25

30

35

40

45

50

55

CGCGCTCCGCCTCGTCGCCGAGCGGGGCCGCTGATGCAACAGATGCCGCCCGGCGCGATGCT
GGCGGTGCCCCCTCGCGGAGCCCCGAGCTCGCCCCCTACCTCAGCGACGACATCTCGCTCGCGGC
GATCAACGGTCCGGCTCTCTCGGTGGTCGCTGGGCCGATCGAGGCCATCGACCGCTCGCGGC
CGAGCTCTTGGAACACGGGCTCTCGTGCCGGCGACTCCACACGCGGCACGCCTTCCACTCGAA
5 GATGATGGCCCCCGTCGTTGACGCCTTTACCCGATGCGTGTCCGCGGTGAGCGCCGCCGCC
15 GTCAGGCCACTTCTCTCGACCTGACGGGCGGCTGGATCTCCCCGAAGCAGCGACCATCCC
CGCATACTGGGCCCGGCAGCTCGTGGAGCCGGTGGCTTCGCCCAGGCCGTGAGGCAGCTGCT
GTCCGAGTCGACGTGGCTCTGGCTCGAGCTGGGTCCGGGCCAGACCCTGAGCCCGCTCGTACG
GCAGCAGGCCCGCGCGGATGGCGGCCAGGTGGTCTCGCTCGCTGCCGCGCGCAAGGACGC
20 10 GGGCGCCGACCACCTCGCGGTCTCGAGGCGCTCGGCCGTGTCTGGAGCGCTGGTGGGACGGT
CGACTGGAAGCGCTTTCACGAGGGCGAGGCGCGGCGGGTGGTCTGCTACCGACCTACCCCTT
CGAGCGCAACGATACTGGGCCTCTCCGCGCCACACGAGCGCTCCGCCGGAAGCGATAATCAA
GCCGCTCTCGGAAGAACCACAAACGTCGCGGATTGGTTCTTCTCCTCCCTGGCGGCGCTC
25 GGATCCTCCGGTCTCGTTCGACGCGCAGGCGGTGACCACGCGGCGCTCTACGTGGCTCGTCTT
15 CATCGGGGACGAGGGCCTCGGCGCGGCGCTGGTGGAGGGCCTCGCGCGGCGGGGACGAGGT
CGTCGCGGTGGTCACGGGTGAGAGGTTCGAGCAGACGGGCACGAGCGCTACAGATCGATCC
30 CGCCGGAATGGCGATGTTGCGTCCCTCTTCGCGCGGCTCGAAATCGAAGGGCGCATGCCGGA
CCGATCGTCCATGCCTTCTGCACGTGCGCTGCGGACGGCGCGCATCGAGCGCGAGCCGC
GCTGGAGATCGAGCGCAGGCTGGGCTTCGATAGCCTCCTCCTCCTCGCCCAGGTGATCGCCGC
20 ACAAGGCATCCGAAGCCGCTGATGCTCGGCGTGATCACGACCCGGGCGCACTCCGTCATCGG
35 AACCAGATCATCGAGCCCCCTGCGCGCTCTGGTGCTCGGCCCTGCCGCGTCATCCCGAAGA
AATACCCCATGTCTCGTGCCGGAACATCGATATCGATCTCCCGGGCGAAGGCGGGCGCGCGGA
GATCGCGGCGCGCCTGATCGCCGATCTGGAGCGAGAGTCGCCCCACTCGGTGGTGGCCTACCG
40 CGGCGGCGCGCGCTGGGTGAGAGCATAGAGCTCACCGATGTGCGCCGGCGGTGAGCTGGCGC
25 CGCCCCGCGCCTCCGCCAGCGCGGGCGTACCTCATTACCGCGGCCTGGGGGGCATCGGCCT
CGTGCTGCAGAGCTCTTGGCCCCGAGAGGCGCACCGCGCTGATCCTGGTTGGGCGGACAGG
CCTGCCAGCGCGGCGAGGGTGGGACGACTGGCTCGCGGCGCACGGCGGGCGACGCGACGAG
45 CCGAAAGATCCTCCGATCCGCGCGCTCGAGGAGGCGGCGCGGAGGTGAAGATCGCCGCGGC
CGACCTCTCCGATTTCAATGCGATGCGGAGCGTCATCGAGGAGGCCCGACGCGCTTCGGCCG
30 CATCGACGGCGTCATTCACTCCGCCGGCATCGCGAGTGGAGGCATGATCCAGCTCAGGACGCC

5

56

10

15

20

25

30

35

40

25

45

30

50

55

GATGGCGGCTTGGCGCGTGATGGGCGCGAAGGTCGGCGGCACGCTCGTGCTCGATGCGCTCCT
CCGGGACGAGCGTCCCGACTTCCTCCTGATCTGCTCGTTCGCTTGGCCTCGCTGGTCGGCGGCGC
CAGGAGATCGATTACTGCGCCGCCAACGCCCTTCCTCGACGCCCTACGCGCAGAGCCGCGAGGG
CGAGGAGGGATGCCGCGTCATCTCGGTGCAATGGGACACGTGGAGTGACGTCGGGATGGCGGT
5 GGACTTCAAGCTCCCGGCCGATCTCCAAGAGGGGCGCCGCGAGAGCCTGAAGCGGGGCATCAG
CTCGAGCGAGGGCGCCGAGGTGCTCGGCCGATCTTGAGCGCAGGCATGAGCGGCCCGCTGGC
15 GATTTGCACGTGGATCTACCAGCGTACAAGCAGTCTGTACGACACGCCGATCGCAGCACGA
GAAACTCCCGCCGCGCCGCGCCGATGCACTCGCGCCCAACGACCACGGGAGCCTATGTGCTCC
CGAGACCGAGACCGAACGGCGCATCGCCGCGATCTGGCAGGATCTCCTCGGCCTCGAGCAGGT
20 AGGCGCAAACGACGATTTCTCCAGCTGGGCGGCCATTCTGCTGTTGGCCACGAGGTCCTGT
TCGCGTCCTGCAGACCCTCAAGGTGGGATCTCGTTGCCGAGTTCCTTCGATGCGCCGACGGT
CGCAGGGCTTTCGCGCCTGGTCGACGCAGCACGGGCCGAAGGCGCCGACCCGTCCGCGCCGC
AATCGGCCGTGTCGAGCGAGACGCTACCGAATCAAGCCGCCCGCGGCCGAACAGGCCGCCCG
25 CACCAAGCCGTAACAAGAAGGGGATCGAGTCATGGAACCCGTGGCGGGCGTGACATGAATCA
15 GCGCGCAAAGCAGCAGGAGACCTGCGTCTTCCCGACCTCCTTCGCGCAGCGGCGGCTCTGGTT
CCTCGACCAGCTCGAGCCGGGAGCGCCGTCTACAACATGCCCGCCTCCTTCGGACGCGCGG
30 GCGGTACGACGTGACTCGCTCGTGCAGCGTGAACGAGATCGTGCGGCGCCACGAGTCGCT
GCGCACGACCGTCGATGTCATCGATGGCGAACCCGTGCAGGTGATCGCCCCCTCGTGCAT
CGAGGTGCCCCGTGTCGGACCTGAGCGAGATCGACGAGCCGAGCGAGAGCGGAGGCCCGCG
20 GCTCATGGCGGAGGAGAGCCGCCGCCCTTCGATCTCACGCGAGGGCCGCTGCTCCGAGCCAA
35 GCTGCTCCGGCTCGGCGAGGCCGATCACGTGCTGATCTTGACGATGCATCATATCGTCTCCGA
CGGCTGGTCGATGGACGTGCTGTTCAAGGAGCTTCCACGCTCTACGCCGCCCTTCCACGAGGG
CCGCCCGTCGCCGCTCCCGGAGCTGCCGATTCAATACGCCGACTTCGCGGTGTGGCAGCGGGA
40 GCTGCTCCAGGGCGAAGTTCTGGAATCGCACCTCGGGTACTGGAGAGAGCACCTCCGCGGCGC
25 CCCCACGCTGCTGGAGCTTCCGATGGACCGGCCCGGCCGCGCGCAGACGTTCCGGGGCTC
CCAGCGCGCTTCCGACTCCCACTCTCCCTGCAACAGGCGGTGCAGGCGCTCAGCCGGCAGGA
AGGCGCGACCCCCCTTCATGACGCTGCTGACGGCGTTCAGCGTGCTGCTCTCGGTTATGCGCG
45 GCAGAGCGATCTGGTGGTTGGCACGCCCATCGCGAATCGCACCCGAGCAGAGCTGGAGGGGCT
GATCGGCTTCTTCGTCAACATGCTGGCGCTGCGCATCGACCTCGGGGGCGACCCGAGCTTCCG
30 CGAGCTGCTCGGGCGGGTGCGGGAGGTGACGTTGGGCGCCTACGCGCACCAGGACCTGCCCTT

5

57

10

15

20

25

30

35

40

45

50

55

CGAACGGCTGGTGGAGGAGCTGTACCAGGGCGGAGCCCCAGCCACAGCCCCCTTGTTCCAGGT
GTCTTCACGTTGCAGAACACCCCGATGGATGCGACGAACAGAGCAGACATTGCATCGGGTGG
CGGCGCGCTGGTGGAAATGAAGGCGGCGAAATTCGATCTGATCCTGGAGCTCTCGGAATCGCC
GCAAGGGTTGCTCGGCACGTTTCGAGTACAACACCGACCTGTTCGACGCCGGCACCATCGAGCG
5 GATGGCCGGCCACCTGGAGGTGCTGCTCTCCAGCGCCGTGCGGCGCCGGATCGACCCATTGC
15 GGAGCTGCCGCTCATGGGGGCCGAGGAGCGCAGTCGGGTATTGGTGGAGTGGAACTCCACTGC
CGCGCTGTATCCCGAGGACCATTGCATGCACGAGCTGTTTCGAGCAGCAAGTGGAGCGGTGCGC
CGAGGCGACCGCGGTGCTCCTCCAGCAGCAGACGTTGACGTATCGAGAGCTGAACATGCGCGC
CAATCAGCTCGCGCATCACCTGCGGAGCCTGGGCGTGGGCCCAGAGGTGCGCGTGGGTGTGA
20 TCTCGAACGGTCAATCGAGACGGTCTGTGGCGATCCTCGGCGTGCTCAAGGCTGGCGGGGCTA
10 CGTGCCGCTCGATCCGACGTACCCAGCGAGCGCCTCGGGCTCATGATGGCGGACGCAGCGCC
CTCGGTGCTGCTCACGAGGCGTCTGCTCTCTCGAAGCTGCCGCCCCACGGGGATGCAACGCT
GGTACAGCTCGACGCGCTGCACGAAGCGCTCTCCAGGCTGCCACACCATAACCCCGGAGCGG
25 CGTCACCGCCCAGAACCTCGCATAACGTATGTACACTTCCGGCTCGACCGGGCGGCCAAGGG
15 CGTGCTCGTCGAGCACCGCGGCCTCTGCAACCTGCCACCGTGCAGGCCAAGCTCTATGGAAT
CGCGCCGGGGCAGAGGCTCCTCCAGTTCGCGCCGCTCTGCTTCGACACATCGTTCTGCGAGAT
30 CGCGCTCGCGTTGCTCTCGGGAGCGACGCTGGTTCATGGGCACGGCGGACGAGCTTCTCCCGGG
ACCTCCGCTGGTCGAGCTGCTGAAGAAGCACGCGGTACGGCGATGCTCCTGGCCCCCTACCGT
GCTCGCAGCGCTGCCAGAACACAGAGCGCGGCGTTGCCGCTGCGCGTGCTCACGATGGCCGG
20 TGAGGCGTGCCCGCGGAGCTCGTCAAGCGCTGGAAGGCACCCGGACGGCGCCTGTTCAACTC
35 CTATGGCCCGACCGAGACGACCATTGCGCAAGCTCCGCAGCGGACCTGTCCGACGAACGGAT
CCCGCCCATCGGCCGTCCGATTGCCAATACGCAAATCTACGTGCTCGACGAAGCGCTCGAGCC
GGTGCCCATCGGCGTGCCGGGCGAGATCTTCATCGGCGGCGTGGGCGTCCCGGGGATATCA
40 CGGGCGTCCGACCTGACGGCCGAGCGATTCTGACCCGACCCCTTCGGGCAAACCAAAGGGC
25 GCGCCTGTATCGGACCGGCGATCGGGCGCGTGGCTGCCGGACGAAACCTCGAGTTTCTCGG
TCGAAACGACGAGCAGGTGAAGGTCCGCGGTGTCCGCATCGAGCTGGAGGAGATCCGCGCGGC
GTTGCTCAAGCACCCGGCGGTGCTCAAGCCGTGGCCGTGGTGGCGAGGACACGCCGGGGGA
45 CAAGCGGCTCGTCGCGTATGTGCTCGGACGCGGAGGAGCGCGGTGACCGCCGCGAGCTGCG
CCAGTCCGTGAGCGAGCGATTGCCTGCGACCATGGTGCCATCGTCCTTCGTGGCGCTCGACGC
30 CTTGCCCCTGACGCCGAATGGCAAGGTGGACCGCCGCGCGTGCCGGAGCCCAGCAGAGCGC

5

58

10

15

20

25

30

35

40

45

50

55

CGGCGGCGAGGACCACGTCGCGCCCGCAACGCCGTCGAGGAGGAGCTCGCCAGGATCTGGGC
GAGCGTCCTCCGGCTCGAAAGGGTCGGCGTCCACGACAATTCTTCGAGATCGGCGGCGACTC
GATCCTGAGCATCCAGATCGTGGTSCGCGCGCAGCAGGCAGGGCTGCGCCTCACCCCGCGTCA
GATGTTCCAGCACCAGACCATCGCCGAGCTTTCGACCGTGGCTAGAGCCGTCGAGGCGGTCCA
5 CGTCGAGCAGGACCCGGTGACCGGTCCCGCGCGCTCACGCCGGTGACGCGTGGTGGCTGGA
GCAGGAGGCGGCGAGCCGACCACTTCAACCAGTCGATCTTCCTCGAGGTACGCGAGCGGCT
CGACGAGAGCGCGCTGGAGCAGGCCATCGCGCATCTGATCGACCACCACGACGCGTCCGGTT
GCGCCTCGCGCGCGACGAACGCGGCGCCACCAGGTCTTCGCCGCGCGGGAGGCTCGACCCC
ATTTACGCGCGTCGACCTCGGGGCGCTGCCAGCGCCGAGCAGATCTCCGCCATGGAGAAGGC
20 CGCGAGCGAGGCGCAGCGAGCCTCGATCTGGCCGCGGGCCCGGTCTCCGCGCCGTGCTCTT
CGACCTCGGCGAGGTGCCCCGCAACGCGTGTCTCGTCATCGCCACCATATTGCGGTGACAG
CGTCTCCTGGCGGATCCTGCTCGACGATCTCTTTGGGGCTATGAGCAGGCGCGCGCGCGGA
GGCCGTACGCCTGCCGCCAAGACCAGTCGGTCAAGCGCTGGGCGGAGCTGCTACCGAGCA
CGCCGGCTCCGAGGCCGTCAAGGCGGAGCTCGGCTACTGGCTCGACTCATCGCGACGAACGGT
15 AGCTCCGCTGCCCGTGGATCGACGGGCGGCGAGGACGTGTGGGGCTCGGCGGCCACATCGT
CGTCTCGCTCACGCCGAGCAGACGGAGCAGCTCTGCGCGAGGTGCCGAGGCGTACCGCAC
ACGGATCGACGACGCGCTCCTCACTGCGTTTCGCGCAGGCCATCGCTCGGTGGACGGGCTCGCC
GGCGGTGCTCCTCGACCTCGAGGGTCACGGGCGCGAGGAGCTCGCCGGCGTAGACCTCACGCG
CACGGTCGGCTGGTTTACGGCCATGTACCCGATCTACTCCGCGTCGACGCGGCGGATCCGGG
20 TGAGGCGCTCAAATCGATCAAGGAGCAGCTCCGCGCCGTGCCAGGCCGCGGGCTCGGCTACGG
CTTGTTGCGTTACCTTCGGTCCGATACCATCGCCGAGGTCCGCGCGTTGCCGAGGCCGAGCT
CTGCTTCAACTACCTCGGCCAGCTCGATCAGGCGATCCCCGAGGCTGCACCGTTCCGGCCGGC
GCGCGAGTATCAAGGCTCGGAGCGCAGCCCCGCGGCCATCGCGCCACCTCATCGAGGTGAA
CGCGAGCATCGCCAATGGGCGCCTGTACGCCACGTGGACGTACAGCGAGCGCGCCACGAGCC
25 CGAAACCATCGAGCGCGTCGCGGCGAGCTTCGTACGGCGCTCCGCGCGCTCATCGCGCACTG
CACCTTGCCCGAGGTGCGCGCAACACGCTTCCGACTTCGACAAGGTGCGCTGCGCCAGGA
GACCATCGATGCTCTCGACGCAATCGACGCGGGCCCCGGGCGTCTGCGAGGGGGAGCCGAAT
CGAAGACGTCTACCCGCTCTCGCCGCTCCAGGAGGGCATCTGTTCCACACGCTCTACGCCAC
CGATTACAGGCGTATGTCGAGCAGTTCCACTGGACGCTGGAGGGCGATTTCGACGCGGAGGC
30 GTTACCCCGCGCCCTCCAGGACGTCTGTCGCTCGGCATGCCGCCCTGCGCACGTGTTGCGCTG

5

59

10

15

20

25

30

35

40

45

50

55

GGAGCGCCTCGATGCTCCACTTCAGATCGTCCGCACGGGCGCGGTCTCCCCGTCGAGCACCA
GGACCTACGCGGCCTCGCCGCGGAGGAGCAGACCGCGCACATCTCCCGTTACGTCGAGGCAGA
GCGCCAGCGCCGGTTTCGATCTGCGAAAGGCGCCCTCATGCGCGCCGGGCTGCTCCGGCTCCG
CAAGGACGCCTGGTGCCTCGTCGAGACCATCCACCACCTGATCCTGGACGGCTGGTCGACACA
5 AATCTTGCTCAAAGAAGTGTTACAGCTCTACGAGGCGCACCGCGGACACCGTGCGCATCTCGC
15 GCTGGAGCTCGAGCAGCCGCGGCCCTACGGCGATTACATCGGCTGGCTCGCGAAGCAGGACCA
GGTGCACACCGCGGCCTTCTGGCGSCGCGAGCTCGAGGGCTTCTCCGCGCCGACGCCGCTCGG
CGTCGACCGCGCTGTGCCGCACGACGACGGCGGCCCGCGGTTTGTTGGCGCCGCATCGCCCT
CTCGGGCGACGACGCGGCCCGGCTCGCCGCCTTCGCGCGTCAGCATCAGCTCAGCATGAGCAC
20 GCTGGTGCAAGGCGCGTGCGCGCTGCTCTTGTCACGCTACAGCGGCGATCCCGACGTGCTCTT
10 CGGTATGACCGTCTCGGSCCGCTCGGCGCCGATTCCCGGTATCGAGCGCATGACCGGCCTCTT
CATCAACACCATTCGGTGCGCGTGCGCGAGCCTGCCGACGCGTCGGTGCTCGCGTGGCTCAA
GGC3CTCCAGGAGCACGAGGACAGAGTGCTCGAGCAGGACACAGCCCGCTGGTCGAGGTCCA
GGCCCATAGCGACGTGCCGCGCGGACCCCGCTCTTCGAGAGCCTCGTCGTGTTTCGAGAACTA
15 CCCGGTGACAGTCATCTTCGAGGCCCTCCGGTCGAGGGGCCGACGCGCGGAGGAGGCCT
CCGCATGATCGATGCGCAGTATATCAGTGATCCACCGTATCCGCTGACGGTCGTGCGGCCCTT
CCATGGGACGCTTTATCTCAATATTGGCTACGAGCGCCGCGGTTTCGACGACCAGGCCGTGCA
ACGGATGATCGGGCAGTCACGACGCTGCTCCGGGGCTTCGTGCGAGGGCCGAGACGTGCGT
CCGCGATCTGCCGTTGCTGACGGCCGAGGAGGAGCGCACCCAGCTCCACGCGTGGAATGCCAC
20 GGCCGCGCCGTATCCCGAGGGCCATTGCATGCACGAGCTGTTTCGAGCAGCAAGTGGAGCGGTC
35 GCCCAGGCGACCGCGGTGCTCCTCCAGCAGCAGACGTTGACGTATCGAGAGCTGAACATACG
CGCCAATCAGCTCGCGCATCACCTGCGGAGCCTCGGCGTGGGCCCAGAAAGTGCGCGTGGGCTT
GTGTCTCGAACGGTCGATCGAGACGGTCGTGGCGATCCTCGGCGTGCTCAAGGCAGGCGGGGT
40 CTACGTGCCGCTCGACCCGACGTACCCAGCGAGCGCCTCGGGCTCATGATGGAGGACGCGGC
25 GCCCTCGGTGCTGCTCACGACAGCTGCTCCTCTCGAAGCTGCCGCCCCACGGGGATGCAAC
GCT33TACAGCTCGACGCGCTGCACGAAGCGCTCTCCAGGCTGCCACACCATACCCCGCGGAG
CGG33TCACGSCCAGAACCTCGCATACGTATGTACACTTCGGCTCGACCGGGCGGCCCAA
45 GGGCGTGCTCGTCGAGCACCGCGGCCTGTGCAATCTGCCACCGTGACGGCCAAGCTCTATGC
AATCGCGCCGAGCGACCGGCTCCTCCAGTTCGCGCCGCTCTGCTTCGACACATCGTTCTGCGA
50 GATCGCGCTCGCGTTGCTCTCGGGAGCGACGCTGGTGATGGGCACGGCGGACGAGCTCCTCCC

5

60

10

15

20

25

30

35

40

45

50

55

GGGACCTCCGCTGGTCGAGCTGCTGAAAAAGCACGCGGTACGGCGATGCTCCTGGCCCCCTTC
GGTGTCTCGCAGCGCTGCCAGAACACAGAGCGCGCGTTGCCGCTGCGCGTGCTCGCGATGGC
CGGCGAGGCGTGCCCGCGGAGCTCGTCAAGCGCTGGAAGGCACCCGGACGGCGCCTGTTCAA
CTCCTATGGCCCCGACCGAGACCACCATTTGGGCAAGCTCCGCAGCGGACCTGTCCGACGAACG
5 GATCCCGCCCATCGGCCGTCCGATTGCCAATACGCAAATCTACGTGCTCGACGAAGCGCTCGA
GCCGGTGCCCATCGGCGTGCCGGGCGAGATCTTCATCGGCGGCGTGGGCGTCCCCGGGGATA
15 TCACGGGCGGCGGACCTGACGGCCGAGCGATTTCGTACCCGACCCCTTCGGGCAAACCAAAGG
GGCGCGCTGTATCGGACCGGCGATCGGGCGCGTGGCTGCCGACGGCAACCTCGAGTTTCT
CGGTGAAACGACGAGCAGGTGAAGTCCGCGGTATCCGCATCGAGCTGGAGGAGATCCGCGC
20 GGCGTTGCTGAAGCACCCGGCGGTGCTCAAGCCGTGGCCGTGGTGC GCGAGGACGCCCGGG
GGACAAGCGGCTCGTCGCGTATGTCTCGGACGCGGAGGAGCGCGCTGACCGCCGCGGAGCT
GCGCCAGTCCGTGAGCGAGCGATTGCCCGCGACCATGGTGCCGTCTCCTTCGTGGCGCTCGA
25 CGCCCTGCCCTCACGCCGAACGGCAAGGTGGACCGCCGCGCGTGGCGAGCCCGAGCGGAG
CGCCGGCGGCGAGGACCACGTGCAACCGCGCAACGCCATCGAGGAGGAGCTCACACGAATCTG
15 GGCCGACGTACTTGGGGCAAAGCGGGTGGTGTGCACGACAATTCTTCGATCTCGGCGGCCA
TTCCCTGCTGCTCGTCCGGGTGCATGATCGGCTCGGCCAGCGGTTTCGATCGGCCGCCCTCGAT
30 GGTGCACCTCTTACCTATCCGACCGTGGCGTCTGCTCGCGCGGTTTCCTTGGCGAACGGGCGAA
CGGCAAGCAATCGCCGAGGGAGGCCGCGCGGACGTACGGAGCGCGGCCGCGCGCTGGA
GGCGCGGGCGCGGGCGGCGAAGGCCATCCGTGGCCCGACCTGACCCGGGCACCTTCCAAGCC
20 CCGCCGTTCTCGCACATCCGCCGCTTCGAGCGCGCGTCCAGCGCCGCCGTTTCGCCGACGAG
35 GAGGCGCGAGACGACGGTCCAAGGCCCTTCGTGGGCTCTTTGCCCGCAATCCGGAAGCTGCGC
GGCAGTTCTGTCGCCCTGCAATGCTGCCATTGTAGAGCTCCTCCGCTCGCCGCGGCCTCTTTT
CTTGGCGGCCGTCGCGATTGACCTCACATCCTGATCCCTTCTTGCGTCTGTCAGAAAGTGAT
40 TGACGGCCAGCGCGCGCTTGAGATCTTCCGGCGCGCGGCGATTTCATCGCTCCGGCGCGCG
25 TGA CTGTACCTGCGAAGGGATTATAATGAAACATAACATTGGGTGGCTTCTACCCGCCGCC
TCGCGACGCTTGCTTCTGTCGCCGCTGACGCCGAATCACGGTGAGGATGCGCCCTCCGTGA
CGTCAGCAGAGAGCGGCGCGCGCGGAGCGCTGACTGCGTCTGCGCTCGGGGCGAAGCTCCAGG
45 CGGCGCTGGACGGCGCGCGCGCGGCGGCGGAGCTCCGGGAGCCGAGCGGCGGTCCAGAGCG
GGGACTGTGTCTGGCGGGGCGCCACGGGCGTCTCGGACCTGGTCTGCGAGCACGCCGACGAAGC
30 CTGAGATCTCTTTCGGATCGGCAGCATCACCAAGACCTTCGTCTCTACGCTGATACTCATGC

5

61

10

15

20

25

30

35

40

45

50

55

TCCGGGCAGAAAGCCGGTTGTCGCTCGACGACGCGGTGTCGAAGTATGTGAAGGGCATCCCCG
CCGGCGACCAGATGACGCTGCCCCAGATCCTCGGTACACGAGCGGGCTCTTCGATTACACGT
ACAGCCCCGGCGCTCGGCCAAATGATCGAGGTGGATCCGACCCGCGCCTTCGCGCCGGCAGAGC
TCATCGCCCTCGCCACGGCCGAGGCGCCGTATTTGCGCGCGGGCGCGGGTTTTCGCTATTCTGA
5 ACACCAATTACATCGTGGCCGCGCTGGTGGCCGAGGCGGTGTCGGGCGGGACGCTCGCCGGGC
TGCTCCGCACGCGCATCTAGACCTGTGGCCTCGCGCACACGTATCTGGACGGCGCCGAGC
CGCCGGTCCAAGGGCTCATCCGCGSCTACGGCGACTACGGCGCGGGCTTGGTCGACATCACCG
ACCAGCTGTGCGCCACCGAGGCGTGGGCCGCCGCGCCCTGGTGTGGAACGTGATGACCTCA
ATCGCTTCTTTGCCCTGCTCATCAGCCACGAGCTGCTCTCGTCGACGAGCTTCAGGACATGA
10 CCACCTGGACCCCGACGATGTGGCCCCACGAGCCCGGATATGGCCTCGGCCTCATCGAGCGCG
ATTCTGCGCTCGGCTCCCTCAACGGGCACTGCGGAATCATCTGGGGCTTCAATCGGCGTCGT
ACGGGGTGCCCGGCCGCGCGACGCGATCACCGCGCTCATCAACGGAGCGACGGCGACGCGAG
CGCGGGCTCGTCGACGAGCTCGCGAAGGTGCTGAAAGAGCGCTGATCGAGGCGGAATGGGAGCG
CTTCGGCGGGTGGTGTATGGCGCCCGCGCTCAGAACGCGACGCGCAGCCCCGCGCTCAGCGGG
15 CCTGCGCCGGGGACGCGGCCACGGCGCCCGGACCGACGAGGAGCCGCGGACGGCGGGCGCG
CTCGGCGCGTCTGCTCGCCGACCCCGCGCTTGCCGAACACGTAGAGCGGCAGGCCGACGGCG
ACCCCGGCCACCCCGCCGAGCGCGGTGGCGATCGCCACCTCGGACGCCTCGGCGCGCGCGGCG
CTGCTGTGCTGCTGGCTCGCGAAGACCAGCACCGCGCCGCTGAGGATGGCGGGCGCCGCCAGG
GTCTGTAGGACGAGCCCCGAGATCACCATGACCGGGCTGTTCCACTCCGTGCTCCGCTCCTCG
20 AAGTCGCGGAACGCCGCCCTCGCCGCCGCGAGCTCCAGCTCGATCCGGCGCTGCTCGGCGCGC
CGCTCGTCCGCGGAGCCGATCTCGTGGACGCGCGGGCCTGGCCCTCCAGCGCCCGATGCGC
GCCTCGTGCGCGGCGGGTCTCTCTCCACGTGGCCCTGGCGGGACCCCCGCCACGGCCGGC
GCGACAGAGGGCGCCGACGCCGGGTGAGGCGGGCGCCGCGGGCGGCTCCGCGGCCACGGAA
GGCGCCGCCCGCGGGAGGCGCGGCGGCTCCGCGGCCACGGAAGGCGCCGCCCGCGCGAA
40 GGCGCGGGCGACTCCGCGGCCACGGAAGGCGCGGCCGCCGCGGAGGCGCGGGCGGCTCCGCT
25 GCAACGGAGAGCGCGGGCGCGCGGCGAGCCAGGCCAGCGCCACGCGACGACACGACGGCGC
GCCGCAACCGCGCGGGCGCGCGAAGCGGAGGTGGACCTGCTCCATGCGCGCAGCGTCGCCC
CTCGACAGGGCCGGGTCAAGGCGCGGAGTCCGAGAGCACGAGACCTCCGCGCCGAGGAAAC
45 AGGCGCGCCGGCGGCCCGCGCGGCGGCTCGCCGCTCAACCTCGCGCGGCGGGCGGGCGCC
30 GCCTCCCCCTCCCCGGCGGGCCGCGCTCGGCGGCCACGCGGAGCAGCTCCTGGAAGTGCCGCT

5

62

10

5

15

20

10

25

15

30

35

40

25

45

30

50

55

CCACCGGGCCGAGGTGCGATGCCGTCCATGAACGACGTGAACGCGAAGTACGGCAGCAGCGTCT
GCCAGGCGCGCAGCCAGCCCGGGAGGTAGCGCGGCCGCTCCAGCACGCCCGGGCGCGCAGCT
CCGCCAGCACCGCGATGTCGTGAGCGCGATCCGGCCGCACACGAGCAGCTCCACCTCGTCGC
GCAGCCCGCCCGCAAGGACCGCGGCCAGGAAGGCGTAGACCTCGAGCAGCCCGGCCAGGTAGC
AGGCGTCTTGGTGAACGGCGCGCCGCCCTCGACGAGCCCGCCGCGGCACACGCGCTGGGCGT
CGAAGTAGGCGTCGCGGCGCTCGGCGCCGCGCTCGCGCAGGTGCCGGTACAGGTGAGGAAGC
TCGCGCCCTGCTCGGCCATGTCCACGAGCCGCACCCGCTCGGCGAGCCGGGTAGGCGGCCGA
TGGAGAGCGAGCGGCTGTAGAGCTCGSCGAAGATGGCGAGCCCCCTCTGCGTGCCTGGTGC
GGGGGCCGCCGAGCGCAGGAACGCGCACCGCGGCTGCGCCGCGCGTTGTGCGCGGTGAGCG
CGTGCGTCTCGACCTCGTGGTCCACAGCCCCCTCGGCCTCCACGCCGGAAGGTGCGCTCCG
GCCGATGCGGACCCGGCTCATGCCCAGCACCTTGGCCGTGACGCGCGGCTCGACGGTGA
TCTCGAGGTGAGCCGCGCGCCCGCGGCCACGCGGGCGGCGAGCATGTCCCGAGCGCGC
CGGCGTTCGAGCGGCTCCTCCTCGGATCGCTGGCCTCGTCCCAGCCGTGGACGCGAGGCGCT
CGGTGAGGTGCTCGGCGAGGTGATGTTCTGAGCGAGCCGCCGAAGAACCAGCGAGCGCGCG
CGCCGTAGAGCTCCTGCGACCGCGCGGAGAACGCGCGGCTGCCCGCGGCTCGAGCAGCTCCG
CGGCTTGGATCTGCGCGCGGACGTTGTCCCGCAGCCAGCCGAGCGCCGCGCGTCCCCGTGA
TGGCCCCGAGGAGCTCGCGCAGCTCGGCGACGCGCCGCGGAGGCCGTTCGCGATCGACGCGGT
ACTCGACCTCGGGGAGGCGGTCTCGCCGGCGGCGAAGAAGCGCTCCTCCACCTCGCGCGGCC
AGGCGATGTCTCGAGCAGCTTGAGGGCCTTGCCCTCCGCCAGGCGGCCGCCACCCGATCGA
GCTGCTCCAGCACGGCGCGGTGATGCTCATCGAGCGCAGGATCGCCGAACCGCGAGACGCC
GGAACCGTCATTCCCTCGACGAGGCAACGATTGCCATGTTCCGTGCTTTTTGGAGCGCCGTC
GTCGCGCTCGCCTGCGGGCTCCGGCGATCCAGCGCGGTTGCATGCAGCGAGGTTGTTCCGGG
CTGGCTCGAGAGCGTCTTTGGCCCCACCCGAGACACGAATGCTCCGCGCCGAGCGCGGTTG
ACCGTGGACCCGCCGAGAGCCGATGATACGGTCCGGCCGATGTTCGAGAGTGTAGCTCAACT
CGAAGAACCAGCGCGGCGCTCACCGGACACTGCTACCGGATGCTGGGTTCCGTGGTTCGACCC
CGACGACGCCGTCCAGGAGACGATGCTGCGCGCCTGGCGGAGCCTGGATAAGTTTCAGCGGCG
CTCGTTCGTGCGCACCTGGCTGTACCGCATCGGACGAACGTCTGCATCGACCTGCGGGCCGA
CCGCGCGCGCGGGCGCGCCCCATCGAGGAAGGCCCGGTGCGCACGGTGGACGACGCGCTCGA
GACGCGCCCGCGCACCCACTGGCTGGAGCCCGTCCCCGACGCGCACGCCCTGCCGGCGGACAT
CGACGCCCGGAGCGGGCGATGCTCCGCCAGAGCATCCGCTCGGTTCTGTCGCGGCGCTCCA

5

63

10

15

20

25

30

35

40

45

50

55

GCACCTGCCGCCGAAGCAGCGCGCCGCGCTGCTGCTCACGGAGGTGCTCGGCTGGTCCGCCGC
GGAGGTCGCGGACAGCCTCAACACCTCGGTGCGCGGATCAACAGCGCGCTCCAGCGCGCGCG
GGGACGCTCGCGAGCCGCGATCTCGGGCAGCGCGCCCCCTCGCTGCCGGAGCCGAGTCCGC
GCTGCTCGACCGCTACGTCAACGCCTTCGAGCGGTACGACGTGACGCGCTCACGGCGCTGCT
5 GCACCAGGACGCGACCCTGTGATGCCGCCGTTTACCCTGTGGCTCCGCGGCCACGAGTCGAT
CCGCGCCTGGCTCGTGGGCCCCGGAGCGGGCTGCCGCGGGTCGCGGCTCATCCCGACGGCGGC
GAGCGGCTCGCCCGCGTTTCGCGCAGTATCGCCCGGCGCCGGAGGGCGGCCACCGGGCTGGGC
GCTCATCGTCCTCGACGTGCGGGGGACCGCATCGTCAGCATGACGTCCTTCCTCGACACCGA
GACGCTCTTCCCGCGGTTTCGGCCTGCCGCTCGATCTACCGGCGTAGCCGCGGGCGCCCTGCCT
10 GCCTCGCCGCGGGTGCCCTGCCTGCCTAGCCGCGGGCGCCCGGCTGGCCACGGGCGCCCGGC
CTGSCACGGGCGCCCGGCCAGCGACGGGGCGACGATTTTTTTCTGAGCGACCGATGAGTCCT
GACGGGGCCGGGGTCTACGGGGTGAATCCAACACGGAGGCACCCATGACCGTGACCATCGC
CAGCATCGATCATCGTGACCAGGACCTCATGACCGGGCCCCAGGCCAAGGCGCCGGCCCGCGC
GGCGCGCCCGACGCGGCGCCGTCCAGGCGAGCCGTGTGGGCGGGCCGCGTCTGAGCGGGCT
15 GGCCACGCTGTTCTTGACGTTTCGACGCCGCGGTGAAGGTGCTGAAGCTGTTCCCGCGGAGGC
GTGACCGCGAGCTCGGGTTCGGGCGCACCTCGTCCCCACCCTCGGCTACCTCCAGATCGC
TTGCCTCGTGCCCTACCTGATCCCGCGCACCGCGGTGCTCGGCGGATCCTGTGGACCGGCTA
CCTGGGCGGCGGATCGCGATCCACGTGCGGGTCGAGAACCCTCTTTCAGCCACAGCTCTT
CCCCATCTACGTGCGCGCGTTCCTCTGGGCGGGGCTCTGGCTGCGCGACCGCCGCGTGCGCGC
20 GCTGACCGCGAGCCCGTCTGTCGAGGGCCGATGAGCTTCACGTTTCACGAGATCCATCACGG
TAAAAGGAGAAGCGAGCCATGACCACAAAGAACCCCCGCAAGCTCTTCGTCAACCTGTCCGTC
CGCGACCTGAAGCGATCGATGGAGTTCTTCAGCAAGCTCGGGTTCGAGTTCAACCCGCGATT
ACCGACGAGAAGGCCGCTGCATGGTCTGTCAGCGAGGAGGCTATGTATGCTCCTCGTGGAG
TCGTTCTTCAAGACGTTTCATGAAGAAGGAGATCTGACGACGAGCACGACACGGAAGGGCTC
25 TTCGCGCTCTCGTGCAGCAGCCGGGCCGAGGTGACGACATGGTGAAGAAGGCGGTGCGGGCG
GGCGGTGCGACGCGATGGATCCCGAGGATCACGGCTTCATGTACGGGTGGAGCTTCTACGAC
GTGATGCGCCACCACTGGGAGGTGATGTGGATGGATCCCAAGGCGATCCAGCCGTAGCCGACG
45 GGGCTGGGCGCGCCGCTGGAAGAGCCCCGTGAGGCGGGGAGGCGGGAGGATCACCGTCTTC
GTAACCCACAGCGATGCAGTATCCGTGCGCTTCGTATCGAAGCACGGCTGTTACGGGCGCGT
30 CAGAGCGCGTCGACGGTGATGCCGAGCCGACGACGCGACACGGGCACGAGCGTGGCTCCGATG

5

64

10

15

20

25

30

35

40

45

50

55

GAGATGAGCCGAGTCTCGCCCATGGTCTCGGGGTCATGAATGGATGAGTAGGGGACTCGCTCC
TTCGTCACGTCGTGCTCGACGGCGACGGCGAGGCCGAGCTCGAAGTGCACGGGGCCTGGACCG
AAGATCCAGCTCGCCCCGGCGGAGCCCCGACGAAAAGCGTGTGCGCGTCGACGCCAGGGCCG
TCGTCCCAGCCGGGCGATCCCACCGCGGTGTAGGTGTGTTTCCCGAAGGAACCCGCGAGCGAG
5 AGTCGAAGTCCGACCGGCGCTCGCCACGCGACGCCCGCTGTGCGCGCCGACGCCGCCGAAGCTC
15 TCCCCGAAAGGCTTATCCCCTGTCTCGATGAAGCCACCCACCTCGATGACGCTGATGCGGTAC
GTGAGCGCGAGATTGAGGTGCACCCAGCGCTGTCCGAGCCCGAGTAGAGGCCGGGCCCCACC
TGCACGCTGAAATCCATGCTCGGCGCGGATCCGCGCGCAGGAGCGACGCCAGGGGCGCTGCCC
TCCTGCGCGCGGGCCGTCCCGACGCAAAGAAAGAGGGCTGTGCGGAAGAATCCAAGCGAGATC
20 10 GATCGAAGTGAGCGCATGTGCGGCCCTGGAGCATCCGCTGTACCAGGTGCGTCGTATTTCATGC
GGCGCGCCGCCGGCGCGCCGCGCTGGCCTGTCCGACGCGAGATCACGAATCCGCCATCGCT
CCCCTGGGCCGCCGGCGCTCTGGTTCGCTGCGGGCGTGCGCCGGCGCTCGTGTGGCCCATG
GCAACCTTGTGCGGTGTGCTCGAACAGCACAGAGAGTATCGCGTCCGCAACAACCGCGCGA
25 CCCGGCGAGACGCTCGTGGGGCCCCCTGCCTCCCCACTTCATCATAACGCCATCAGGAGCACT
15 CGACATTTTCATTTCTTCACCTCCACTGGCTGAGGGCGACGGTGTGCTCATCGGCCGGTTGCT
CTGGCGGTTGCTCTGGCGGGGTTTCTGACGCCCCGAACTAACGCTTCGAGCGCTCCCCCTTGC
30 TCTCCCGTTCTTTCAGCTCCTCCAGCAGGTGCTCGAGGCGCTCGTAGCTGCCTTCCCAGAAGC
GGCGGTAGTTGTGAGCCAGCCGCTGGCGTCTCGAGCGGCTTGGCCTCGATCCGACAAGGCC
TCCGCTGCGCGTCTCGGGCCGCGGAGATCAGGCCCGCTCGCTCCAGCACCTTGAGGTGCTTGG
20 20 AGATCGCGGGCTGGCTCATCGGAACGGCTTCGCCAGCTCGGTACCGACGCCTCGCCGACG
35 CGAGGCGCGGAGGATCGCTCGCCGTGTGCGATCGGCGAGCGCAGCGAACGTTGCGTTCGAGGC
GCTCGGACGGGGTCATTGCATAAATCCTTGGTATAAAAACCAAGTTAGTTATACAACCTGGGGC
CCGGGCGGTCAAGCCTCCAGGCGATGGCGGTTCCGGCCGGGGGCTCCGCTCGCGGCACGCGCG
40 CCGCGCGGCTACGTGCGCGGCGCGGTGAGCACGTCTGCGAGCGTGGCGCCGACCACGGGCTTG
25 GTCAGGTGCAGGTGCAAGCCGGCCCCGCTGGACCTGGCCTGATCGTCCGGGCCGCCGTAGCCC
GAGAGCGCCACCAGGTAGAGCGCTTCGCCGCCGGGCGCGGCCCGCGCCCGCGCGGACCTCA
TAACCGTCGATGCCGGGCAAGCCGATGTCCACGAAGGCCACCTCGGGGCGCAGCTCCAGAAGC
45 TTCTTCACGCCCTCCAGCCCGTCCACCGCCACCGTCACCTCGTGCCCCAGCGCCTCGATGTAC
GCCCCGATCACCCGGCGCACGTCTCCGCGTCTCCACGACGAGCACCCGGCGCGGTCAGCC
30 GCCGCTCGGGCGCCTCGGCGCGCTGCGCCGGAGGCGGCGGGCTCGTTCGCGCTGCGCCGGA

5

65

10

15

20

25

30

35

40

45

50

55

GGCGGCCCCCTCGCGGGCGGGGGCGCCCCGGCGCTCGGGGCAGGCTGCGGCGCCGCCCCGGGG
CCGAGCGGCAGGCGCACGGTGAACCTCGCTGCCCTGGCCCGGCCCGGCGCTCGCCGCGGCCACG
CTGCCGCCGTGCAGTTCCAGGAGCCGCCGACCAGCGTGAGCCCGAGCCCCAGCCCGCCCGTG
CTCCGGTTCGATGGTCTGGTCGACCTCGGTGAACAGATCGAACACCTTCTCGAGCATCGCCGCC
5 GGGATGCCGCGGCCCGTGTGCGCGACCCGAGCACGGCCTCGGGCGCGCCGACCGCCGCCTCG
CGCGTGAGGCGCACCGAGATCGAGCCCCCGCGGGGTGTAATTGCGGGCGTTGGTCAGGAGG
15 TTCGTACCACTTGCTCCAGCCGCTCGCGTCGGCCCCGATGCCGAAGTCCCCGGGCCCCACC
GACAGCGACAGTCATGGCGCCGGGCTCGACGGCCGGCCTACCGCGCGCGCGCGCTGTC
ACCACCGCCGCGAGATCGAGCTCTCGAGGCGCAGCTCCACCGTGCCCGCGTGATGCGCGAC
20 ACGTCGAGCAGATCGTCGACCAGCCGACGAGGTGGCCCATCTGCCGCCGCGCGATCTCCCGG
TAGCGCGCCGACGCGGGCCCGTCCGCTCGCGTCTGTCGAGCAGCGTCAGCGACAGGCTGATC
GAGGCCATCGGGTTCCGGAGCTCGTGGCGGAGCATCGCGAGGAACCTCGTCTTGCGCTGATCG
25 GCGAGCTTCAGCGCCTCGACGAGCCCTCCACGCGCCTCCGGGCGCGCACCTGGTTCGGTCACG
TCGAACGCGAACACGAAGACGCCCTCGACCGCCCGTCCGCGATCGCGCATCGGCTGGTAGACG
15 AAGTTGAAGAACACCTCCTCCGTCTGCGCTCGCCCCGGCGATCGAGCCGACCGGGAGCTCC
TTGCCGACGATGGGCTCGCCGGTGGGACACCGCGTCGAGGAGCTCCAGATGCCCTGTCCC
30 TCGAGCTCGGGGAGGGCGGCCCGGATGGGCTCGCCACGAGCGATCGACCGCCGACGAGCCGC
TGGTAGAGCGGGTTGACCACCTCGAAGACGTGCTCCGGCCCGCGGAGGATGGCGATGGGCCCC
GGGGCTGCGATGAAGAGGTCTTCAGGTACTGGCGCTGCCCCCTCGGCCTCGCGCCGGCGCGC
20 GCGAGCTCGACGTGGATGCGGACCCGCGCGAGGAGCTCCTTCGCGGAGAACGGCTTCACGAGG
35 AAGTCGTCGGCGCCGCGCTCGAGGTGTCGACGCGCGCCTCCTCGCCCGCGCGCGGAGAGC
ATCACACGGCGACGCCGCGGGTGGATCGTCGGCGCGCAGCGCCTGAGCAGGCCGAAGCCG
TCGAGCCGCGGCATCATCAGTCGCTGAGCACGAGATCCGGCGGGTGGGCGGGGCGCGCTCC
40 AGGGCGGCCCGACCGTCGGCCACGSCCTCCACCGTCCACCCCTCCGCCACGAGCAGCCGAGC
25 GCGTACTCGCGCATGTCCGCTTGTGTCGCGGACGAGGACGCGCCCCGGCAGCCTCCCGGCC
GGCCCCCTCGCCCCCGGGCCGGACCCCGGCGCCTGCTCGCCGCGAGCCACTGCGCGGCCTCG
TCGAGGAAGGGCGCGGCGTCCCGCCCCCGCGGGCCGGCGCCGAGGCCGGCGCGAC

or its complementary strand,

30

5

66

10

(b) DNA-sequences which hybridise under stringent conditions to regions of DNA-sequences according to (a) encoding proteins or to fragments of said DNA-sequences,

15

5 (c) DNA-sequences which hybridise to the DNA-sequences according to (a) and (b) because of a degeneration of the genetic code,

20

(d) allele variations and mutants resulting by substitution, insertion or deletion of nucleotides or inversion of nucleotide segments of DNA-sequences according to (a) to (c), wherein the variations and mutants offer isofunctional expression products.

25

7. DNA sequence according to claims 1 to 5, wherein the DNA is selected from the group consisting of

30

(a) the following DNA Sequence:

Seq ID No 2 (>pEPOcos6 region)

35

20 GGATCACCTGCGGCGGATCGCCGACCTCGTGCTGGTGTTCGGCTCGCTGGATGAGAAGCCGG
CGGCGCTACTGATAGAGACGGCGACGCCCCGGGCTGCGGTGGAGCGGTTGCGGGAGATGCTCG
GCTTTCGGGCGGCCACCTGGCGAAGCTGTCTTCGACGGTTGCGAGGTCCCCGAGGCTCAGC
TGATTGGCCGGCCCCGGCTTTGCGCTGATGTATCTGGCCCCCTACGCCCTGGATTTCGGTTCGGG
TCAGCGTCGCCTGGGCCTGCCTGGGCATGATCCGCGCTTGCTGGAGACCTGCGCACAGCACA
40 25 TCCTCACCCGCCGCACCTTCGGCCACCTGCTAGCCGATCACGGCATGATCCAAACCTGATCA
CCAACCTGGGGATTACCAACAGGCGACGCTGCTCCACACGCTGCAGGCCTGCCGCGCCAGGG
ATCGCGGCGACGTGACCGCTCCGAGGCCACCCTCGCCGCCAAATACCTCGCGTCGCGGACGG
45 CGGTCCAGGAGACGACCAACGCGGTCCAGATCATGGGCGCGCTGGGCTGCGACGAGGAGGGCG
CGATCGCCCGCCACTTCCGCGACGCCAAGACGACCGAAATCATCGAAGGCAGCAACCAGATCA
30 TCGAGGCGCTGCTGGCCAAGAACATCGCCCGCGCCGGTCGCGACAACATATCGCCGCTTCCTCG

50

55

5

67

10

15

20

25

30

35

40

45

50

55

ATGCGGAAGTCGAGCCCGGTGCGGCCGGAGGCGCACCATGACGAGCGCGGTCCCGACGCGTCA
AACCAGCCTGCTCGACGACTTCGAGCGCGTCCCGACGTCGATCCAGAGCGGATCGCCGTCCA
CGCGAGCGAGACGAGCCTGCGCTATGGCGACATGAATGCGCGCGCCAACCGCATTGCCACGG
GCTACGGGCGCGCGGGATCGGGCCCAATCAAATCGTGGCGGTGGCGATGGCCCGACGCCCCGA
GCTGATGATCGTGCTGTACGGCATCCTCAAGCGCGCGCGCTACATGCCCATCGCCCGCGA
CGCGCCGCGCTGCGCCGCGATCATATGCTGCGCGAGAGCCAGGCTGCTCTGATGATCGCCGA
CGAAGAGATCGCGGGACTCGCGGCCCGGTGCTGACGCGCGCGGACCCGTTCTTCGCGGCCAT
GCCGGACCACAACCCCGAGCCGCGTCACGACCCGACCGACCTGATTTACGTCATCTACACCTC
GGGCTCGACCGGCCAGCCCAAGGGCGTGGCCATGGAGCACCGCGCCGTGTGGAATCGCCTGAC
TTGGATGCAGGCCAGTATCCAATCGACACGACGAGACGTGATCCTCCAAAAGACGCCGATCGT
CTTCGACGTGTGGTCTGGGAGCTGTTCTGGTGGCCGCTGGCCGGCGCCTCGGTGGCCCTGCT
GCCGCAATCCATGGAGAAGTTCCCTGGGCGATATCGGCGACGGTGGCGCGGTGCGGGTGAC
GGTGTATGCATTTCTGACCATCGATGCTGATGGCCTTCCTTCAGGTGGTGGCGGGCGGCCGA
GATGGCGGACCAGATGAAGGGCCTGCGCTACGTCTTCTGCAGCGGCGAGGCCCTGGCGCCGGC
CCACGTGTCAGCCTTTCAGGAGCACATCAACCGAGCGGGCAGCATCAGCTTGACCAACCTCTA
TGGACCCACCGAGGCGCGGTGACGTCAGCTACTTCGACTGCCCGCCCGGCGCGTCACTCGC
GCGGTGCCGATCGGACGAGCGATCACCGGCATCCAGCTGCTGGTCATGCGCGACGGCGTGCC
TCAGCCGCGCGCGTTCGAGGGTGAGCTCGCCATCGGCGCGCTTGGTTTGGCGCGCGGTACAT
CTCACGGCCAGACCTGACCGCCGACCGGTTCTGTGCCGATCCAGGCGGCGACGGCCAGCGGCT
CTACCGCACCGGCGATCTGGTGGCGAGGGACGCGGACGGCGAGCTGGTCTTCCTGGGGCGCAT
CGACCATCAGGTGAAAATTCGCGTCTGCGCATCGAGCCCGGGGAAATCGAGGCCAGATCAG
CGCCCATCCCGATGTGGCCGACTGCGCGCTGATTATCGAGCAGGACTCGGAAACCTGCCCAA
GCTGACCGCTACATTGTCTGTGGCGGACCGGGCTTGACCCGGAAGGCGCTGCTACAGTTCTT
GGGCGCGCGGCTGCCCGACTACATGCTCCCGAACCGCTTCTGACCTCACGGAGCTGCCCCGT
GACCGCCAACGGTAAGCGGACTGGCGCGCGTCTGCTCGGCCCGCTCGAGACCTGCTCTCCC
TTTCTCCTGAATCCAACCAATACGAGGGATTATGTTACACCGATTCCACCGACCGTTTCG
CCCTGAGCCGACCGCTCTTTTCGCGGTACCTCGCGCACGATCCGATCGTGAGGGCGTGCTGG
CGGCGGACCATCCAGGCTGGGTCTTGGTGGACCGCGAGCCCGAGCCGCGACGGCGCTGCTGT
GGGCTTTTCCGATCGGCTCTTCTGCGTGGGCGCAGCTGACACGCTGACCCCGCACGCGCTGG
CCGAGCTGTTCCACGACCGACTGATCCCCAGGCCGTAAGATCGGGCAGCCGTTTTTCAGG

5

68

10

15

20

25

30

35

40

45

50

55

TTCAGGGCGAGACGGTCGACACCTGGTCCGACCACCTGCATCAGGTGTCGCCGCACGCGACAG
TCTCCTTCCGCCAGGCATTCCGCTTCGACCGCGACCTCTTCGAGCGGCTGCCAACCAAGCCGG
AGCTGGCAGAGGCGCGGCTCGTGCCAATCGACGCGCGGCTGCTG3CCGAACAGGCTGATCTGC
GCGAGCGGATACTGGCCTCCTGGTCCAGCGAAGCTGCCTTCCATGCGCGCGGTTTCGGCTTCT
5 GCTACCGCGTAGGTGACCAGCTGCCGAGCGTGTGCCTGGCATCGCACGTAGGCGGCGGCGCGG
15 CCGAGCTGAGCATCAACACCGAGCTCGAAGCGCGCAATCGAGGTATGGCAACGCGGCTGTGCC
GGCGTTTCATCGCCGAATCGCTGCAGCGCGGCTGACGCCTTGCTGGGGCACCGAGACCTTTC
GCCTGCCGTCAATCGCGCTGGCCCAAGCTCGGTTTCATCCCGACCTTCACCTTCCCCACCT
ACTGCTTCGCGACCGGCACCGAACAGCCGACGACAACCTTCCTAGGCGAGCTGTACTACAGGG
20 10 AATCGCGCATCGCCGGAAGTGGGACCGATGAGCCGCAAGCGGTTTCGGCTGGCGCGGGGTGGA
GCCTGGCCGCGGACACCGAGCGTGCCGCGAGCTTCGCCCGCACGCGCCCTGGCCGAAGGGTGGG
CCGGCCACTCGACTCTGGCCACCGATCCGGATTTGCGCCGATTGCGCGCCAGCGCCGCTGGC
CCCGCTCAATGTCCCTTGAAAGGTCACGTGGACTCATGATGTCCCTTGAAAGGTCACACTC
25 CGAGTCATGATGATTTGTCACTCCCACCGCTTCATTTTCTCCACGTTCCCAAGGTCGCGGGC
15 ACAAGCGTCAAGGACGTCTCGGCCAAGAGCTATTCCAGGAGGACCAGGTACGTTCCAGATC
GCTCCCAATCCCCACTACCCACCTGAATGGACTGCGCCTTACGAGGAGCACATTATTGCCGCT
30 GAATTGAAGAGCCAGTTGGCGCCGAAATTTGGGACGATTACTTCAAGTTCGCCTTCGTGCGC
CATCCGCTCGACTGGGCGGTCTCCAATTACTTCTTCTTCTCGCGGACCGCAAAGGCCATCCG
GCCCACGAATTCCTGGAGCGGAAGGGCTTCGCCGTTACCATGGACATGTTTTTCGGAGCGGCC
20 GGGCGCCATCCGCTGGTCCGCGCATGCGCTTCAGCCAATGGGAGTTCTTGTGCGACAGCGAG
35 GGCCGGACGCTGGTGGACTTCGTTGGCAAGTACGAGCGGCTCGAGCAGGACTTCGCCGCCGTG
TGTATCCGCATCGGGCTGACCCCGCCGACTTGCCGTGCCTCAACCAGACTCGCCACCAATCC
TTTACCAGTTACTACGACGAGGCTTTGATGCGCCAAGTCAGCCGCGCGTTAGCTCCGGAATTC
40 GAAATTTTGTATTATGCCTGAGGCGGACCCGTTGCTTCGCCACCGGTGGATTATTCGATAAGT
25 TATTATATTTTCAGTTGATCATGTGAATGTGATCCAGCCAACGAGGAGGATACCTCCGCGTG
CGGCTATGGGGGCGCAGAGGTCACCACTACGTGTAGAAATTTGTGCAACACACCACTAGCTGC
CACCATTGGGAGCTTTGACTTGAAGATGAAAGTGGACAAGCGGAATGTCGACGACATTCTCG
45 GACTCACTCCGACACAGACAGGCATCTTGTACCACTACCTGCTGGACCCGAGGCCGACGCCT
ATTTGAAACAATTGACGCTGCACCTGGAGGGGCGCTCGACGTAGCGCGCTTCGCCGCGCCT
30 GGGAGCGCGTGGTGGCGGCTCACGACCAGCTGCGCGCGTGTTCGCTGGCAAGGGATCGAAC

5

69

10

15

20

25

30

35

40

45

50

55

ACCCGGTG CAGATCAT CCTCAAGCAGCACGTGCCGGACCTGGAGTTGGCGGAGGTCCCGCGCG
ACGCCGATCCGGCAGCCTTCTCTGGC3CAATGGGTTCGCGGCCGACCGGGCGCGCAAGTTCGACT
TCGAGACGGTGCCCTTTTCGCATCGGCCCTCTGCCGGA CTGATACCCAACATCACGTGATGCTGC
TCAGCAATCACCATATCCTGATGGACGGTTGGAGTACGGGCCTGATTCTGCGGGACTTCCTCG
5 CCTGCTACGGCGACTCCGAAAAC TGGCGGCCACGCACCCGAACGCACTTCAAGGCGTTCATCA
AGTGGCACCAGAACCGGCCACGCCGGGCGAGGAGCGATTTTGGCGCGACCTGTTGCGCGATG
CGCCCGACGGCGGCTTTCCCGGCCTGGCGCTCGAAGAAGGCACCCGCCACTCGCTTGACTTCG
GCGCCCGCAGCCGCGCTCTCGACGACCGCTTGACCCAAGGCTTGCGCGACATGGCTCGCGACC
TCGACGTCACCCCTCGCCGCGATGCTCCATACCGCTTGGGGCCTTCTACTCCAGCGCTACCAGA
10 ACAGCTGCGAAGTGATATTCGGGACACCGTTTCCGGCCGCAACGTCGAGCTCGCCGGCCTCG
ACGAGGTGGTCGGCTTGTTTCATCAACACGATTCCGTTCCGCTTCTCGGCCGCGGCCGCGACGA
CGCCCGTCGAGGCCTTCCGTGCGGTACAGCGCAATCTGCTGGCGAGAAGCGAGTTCGAAGCCA
CCCCGCTGGTGGACATCAAGGGCTGAGTGGTCTCGGTCCGGGCGCGGAACTGTTTCGACACCA
TCCTGGTCATCGAGAACTATCCCTTGGACCGCGCTATCTTCGAGAGTGATTCCAGCCTGCGGT
15 TGACCGACCACCAAATCTTCGAGCGACCAATTACGGGTGACCCTGACCATCGAGACCTTCA
GCCGGTTGCACGTGACGCTAGCCCATCGCCGTGACCTGCTGGGCGACGCGGCCGCTGAGCGAA
TGCTAGATCATTTACCGGCCTGCTCCAAGCCATGCTGCGCTTCCCTCACCAGCCGTTTCGCGC
GCCTCGAGATGAAAAGCGAACACGAGGCCACCCGCTCCTGCACCAACTCAACCAAACGCGTC
AGCCGCTGCCGTCCCAATCGGCTTTCCACCAGTTGTTCTTCGAGCAGGCCAGGCCGATGGGG
20 CACGACCGGCGCTGTGGTGGCGGCCACGCGCTGGACCTACGGCCAGCTGCTGGAACGTGCCC
TGCGTCTGGCGGGACGGCTGCAGGAAGCCGGCTTCGCCCGAGGCGATGTGCGCCCGCTCAGCC
TCGGCCCGGTTCCGGATCTGATTCCCGGTTTGCTGGGCCCCGCTGTTTCGCCGGCGGCGCCTACC
TGCCGCTCGATCCCAACCTGCCGGCCAGCGCTCGCGGTTTCATCCTCGACGATGCCGGTTGCC
GCTTCCTGATCAGCGACGCGCCACTCGCGGGGCCACGCCGATCCATCCGGACCTGCCGGCG
25 CCAGCCCCGTTGACGTCAATTTTGCTGTGAGGACGGCGCCGCGCAGCCCGCCTACCTGATCT
ACACCTCGGGCTCCACCGGCCAGCCCAAAGGCGTCTGGGTTAGCCACCGCAACCTGATCAACT
TCCTGACGGGCATGAGCGCAATCCTGCCGGTCGCGGCCGACGACGTGTTCTCTCGCTGACTA
45 CCGTGTCGTTTCGACATTTTCGGGCTCGAGACGTGGTTCCCGCTCAGCCGCGGCTGCACGATCG
TCTTGGGCACGCGCGCCGAGCAGTTGGACCCGGCCGCGGCTGCCAAGGCCATCTCCTGCCATG
30 GCGTCACGGTTTACCAGGCGACGCCATCGCGACTCCA ACTTCAACTGGAGCACCCACATTTG

5

70

10

15

20

25

30

35

40

45

50

55

TCCGCGCCATCGGCTCCCTGACGACCCTGCTGGTAGGCGGCGAACCCCTCCCAGCCGAGCTGC
TGCGGCGCGTACGCGAAGTGACCGATGCGCGTATCTTCAACCTCTACGGTCCCACCGAAACCA
CCATCTGGTCCACAGCCGGGGAGGTCAACGCGGCGGACGTCCCGGATATCGGCCGCCGATCG
CAAATACCGGCGTTTTCCTTCTGGCGCGAGACGGCTCGATCCAGCCGCCGGGCTGGTGGGCG
5 AGTTGTGCATCGCCGGCGAGGGCGTGGCGTTGGGCTACCACCGACGGCCGGACCTGAACCGAG
AACGGTTTCGCGAGATTCCGCCGGGCCCGCTGCCCTTTGCCGGCAAGCTCTACCACACCGGCG
15 ACCTGGCCCGCTGGACCGAAGACGGACGGCTCCTCTGCCTGGGCCGTCTGGACGACCAGCTCA
AAGTGC GCGGCCATCGCGTCGAGCCGGGCGAGATCGAGGCAGTGATGGCGGCCACCCGCGG
TCACGCAGGCGGTGGTCTGTCACGCGGCCGCGCAACGGCGAGCCGGTCTTGGTCGGGTTCTGGA
20 CTGCGGAAGGTGAGCCGATGCCAGAGGAAGCGCTGAGCGCTTACCTGGCCGACCGACTGCCGA
GCTACATGGTACCCGAACGGTGCACTCCTCATGAAGGCCATGCCGCTAACCGCAACGGCAAGA
TCGACCGGCGCGCCCTACCCAATCCCTTCGCCTTGACCGAGTCGACCCGGCAGGCGGCGCCGC
25 GCACCTTGGCCCGCACCGCCGGCGAGCATCGGGTTGCCGAGCTGTGGCAGGCCTTGTTCGAC
GCGAGGCGATCGGCTTGGACGAACCTTTTTCAGGCCGGCGGGAACCTCATTTCGGCTTGATT
15 GGCTTCACGCCAAGCTGGAATCCGCCTTCGGGAAGTCGTTCGGATCACCGATTTGTTCCAGC
ATACCAAGTATTTCGAGCCAGGCAGAAATGCTGAGCGGCTCGTCCGTGAGGCGCCGCTCGCGG
30 GAGCCGTGCCGCAACCCCCGGCCGCCGCCGCCCAAGTTGCCTCCTCGGCAGCTAAATCCCCAG
GGGAGCGCGGCGCGGACGCGACGTCGAGCGGCTGACCGCGCAACCGCCCCAACCCACTTCC
GGCCCATCGCCGTTATCGGCCTCGCCGGCCGATTCCCCGCCGACCCGACCTCGACGCCTTCC
20 TTGAACTGCTCACGGAGGGTGCCTGCGGCATTTCGCTTCTTACGCCAAGCCGAGCTGCGCGACG
35 AGGGTCTCGACGCGAATCGAATCGCGTGTCTAACTATGTCCCGGCCAAAGGTTTCTTCGACC
GGGCCGACCACTTTGATGCCGACTTCTTCGGCATCCCGCCGCGGACGCGAGAAATCACCGATC
CGCAAATTTCGGCTTCTGCTTGAAGTGTGCTGGAACGCGCTGGAGCATGCCGGCTACCCGCCCCG
40 GCGGCGGCGAGATCGGGCTTCTCGCCGGCTCCTCGGCCAACTATCACTGGCTCGAATACGTGG
25 GCATTTCCGAGGAGAGCAGCAATCGATTGCCGTCATGATTCAAAACGAAAAGGACTACCTGG
CCACGCGGATCGCCTACCAGCTCGATTGAAGGGCATTGCCGTACCGTGCAAACGGCCTGCT
CGTCGTGCTGACCGCGGTGAGCTGGCCTGCGATGCGTTACACGCCGGCCGCGTGACCATGG
45 CTTTGGCTGGTGGCGTTGGTCTGACCTATCCGTTGCGCGCCGATACCTGCACGAGGATGGAA
TGATCTTCTCCCCGACGCTCGGTGCCGGGCCCTTCGACGCCAGGCGGCCGGCACGGTCTGCG
30 GCAACGGTCTGGGCATGGTGGTGTGAAACAGCTCGACGCGGCGCTGGCCGACGCGCATGCCA

5

71

10

15

20

25

30

35

40

45

50

55

TCCACGCTGTGATTAAGGGCATCGCGGCCAACACGACGGCGCGGCCAAGATCGGCTACACGG
CGCCCTCGCAGAACGGTCAGGCGCGGTGATCCGCGCCGCCATAGGCTCGCCCAAGTCGCGC
CGGAGACCATCGGCTATGTAGAAGCCCACGGTTCGGGCACGCCGCTGGGCGATCCGATCGAGG
TGGCGGGCCTGACCGAGGCCTTTGACAGCCCGCTCGCGGCTTCTGCGCCTTGGGTTCGGTCA
5 AGTCGAATGTGGGTCAATTTGGATGCGGCAGCGGGCATCGCGGGTTTCATCAAGGCGGTGCTCT
CGCTGTCCCATCGGACCCTGTTGCCAGCCTCCACGTCGACACGCCCAACCCGCAGATCCCGT
TCGCCGACGGTCCGTTCCAGGTCAACACGAGACCCGGCCCTGGCCAGCTGCCGACCATCCCC
GCCGCGCCGGCGTCAGCTCCTTCGGCATCGGCGGCACCAACGTGCACGCCGTCCTGGAAGAGG
CGCCGAGTTGGCCGAGCAGCGGGGCGGCGCGAGCGGCAGCTGTTCTTGGTCTCGGCGC
20 GGAATGCGAGCCGATCTGGAGCGACGACCGCGCGCTGGTCCGCCACCTGGCCGCGCATCCGG
ACCTCGCACCAGATGACGTTGCCTTTACCTTGACGCGGGCCGAAACCGATGACCCACCGTC
GTTTCTTGGTTCGCCGCCGACCTCGCGGAAGCCGCCGCGCTCTGGCCGAGCCCGATCCAGTCA
AATCCGCCCGCGCGCGCGCGGACCGCTGCCAGGTCTGGATGTTCCGCCGTCTCGGCTCTCAAT
25 ACCCCGGCATGTGTGGCGGCCTCTATCGCACCAGCCGGCCTTTCCGCGAGCAAGTCGACCGCT
GTTTCGACCTCCTCGCGCCGCGTTCGGATTTGAAGCCCTCGCTCTTCCCGAGCCCGATCAGG
CCATCGACGCATCAGCCCTCGCGGCCATCGACACCGCCAGATCGCCGTCTTCGTCTGCGAAT
30 ACGCGCTCGCACGGATGCTGGAAGGCTGGGGGCTGCGTCCGGATCGGCTGATCGGTTACAGTT
TCGGCGAATACGTGGCCGCCTGCCTGGCCGCGCTTCTCTCCCTGCCCGACGCCCTTGGCAATCG
TCCGCGAGCGTGGCCGATCCTGGCGGCGGCCGAGCCGGGCGCGATGGTCAGCGTGCCCCCTTC
20 CGGCCGAGCGCGTTCGCTCGCTGCTGGAGCCGCGCTTGCCTTGGCCATTGACAACGGCCCCCT
CATGCGTGGTGTCCGGGCCGCTCGAACCGGTGCGCACCTTCACCGCTCGCATGAAGCGGGACC
GGGTCTGGGTGACGCCGCTCCAGGCCGAGCGCCCGATGCATTCCGCCGTGATGGCCGAGGCCG
GCGGCTCACTGCGCGCCATGTTGGCCGGGTTCGCCTGAATGCGCCGGAATCCCGATCTTAA
40 GCAATGTTACAGGAACCTACCTAACCGACGAGCAGGCCCGAGACCCGATTACTGGGCCCCGTC
25 ACCTGTGCGGCAACGTTGCTTCGCCGACGGTGTGCGAACCTTGTGGCCGAGCGCGATCCGG
TGTTCTTGAATTCGGGCCGGGCCGCGATCTGAGCTCCTTGGTGCGCCACCAGATGCCGGAAG
GCGCCGACGAGCCGATCGCACTGATCCGTATCGCGAAGATCCGGTGCGCGACGAAGACCTCC
45 TGCTCGATGGCTTGGGCCGCTGCTTCTGCGTGGGGCGACCCTCCACGGGCAGGCCTTGTACG
CCGGCCGAGGCTGCCGCCGCGTGCCTGCTGCCGTTACCCGTTCCAGGGTCCACGCTGCATGC
30 CGGCCCGCGCCGACTGCCCGCCTGGCGGACCGACCGTGGGAGCGACCACCATCAGCTACC

5

72

10

15

20

25

30

35

40

45

50

55

GACCAGCCTGGAAGCGGGCGCCGCGCTTGGCGGCTGTCGAATCGCTCGCGCCGCAATCCTGGT
TGGTATTTCAGCGACGGCAGCGAATTGGCGGGCGAGCTGGTGGCCGGCCTGCGCGCTTCCCGGTT
GCGCGACCACCTTCGTCTGAAGGTGGGCTGGCGTTTCGCGCGCTTCGCGGGCGGCTTCCGCGCGA
ATCCCCGCGAGGAACAAGATCTCGCACAGCTGTTTCGCGACCCTGTCCGCCGAAGCGATGCTGC
5 CCACCCACATCCTGCACCTGCTCAGCCTGCCGTGCGCGGAGCGCGACTCGCCGCTGGCGCGCC
TGGAGCACCTCACCGAGCTGGGCTTCCACCATCTGCTGGCCCTGGCCCCGCAACTGGAGGCGG
TCGGCGCCCCGAGGTCCGCCTCGCCGTGGTGACAACCGGCCTGGCGGCGATTGGCGGCGAGT
CCGAGCTGCGGCCCGAGGTCCGGCTGTTGCGGGGACCTGTCCGCGTGATTCCCTTTGAATTCC
CGAATTGCGGCTGCGCCTGATCGACCTCGACTCGGCCGATCCCATCTGGCGTAGCGGTTGTG
10 AGCCGTTGCTGCGCGAAATGGGCGCTGCCCCGGGACCTGAAGAAATCGCGCTGCGCGGCACCA
GCCGTTGGGAGTTGGGCTACGAGCCGGTTCGAGGGGGCACCGTGAGCACCATCTCCTCGCGAC
TGCGCGAGGGCGGCTCTATCTGATCACCGGTGGCCTCGGCGGCGCTGGGTCTGGCCTTGGCCC
GTCACCTCGCCCCGAAGTACCGCGCCACCCTGATCCTCGCTGGCCGCGGAGGCGCGCCGGCGC
25 GCGAGCTCTGGCACCAGGCGCCAGCGGAGTTTCGTACCGGTTCGAGCTGCGATCGCACAGATGG
AGGAGTGTGGCGCCCGCGTGATTCCCGTTCGCGCTCGACGTCACCGACGCGGACCAAGTGAACG
CGTTGTTTCGCCACCATAGAAGCTACGGTCGGCAAGATTGAAGGCGTTTTCCACATGGCTGGCA
TCGTTGACGGCGGCATCATTGCAACGCGCACGCGCGCTGCCAGCGACGCCGTGCTGGCGCCCA
AAACGGTCGGAACCTGGATTCTCGATCGGGCTCTCCGCGGCGCCGGTGGCCGCTTCTCTGGTGC
TGTA CTCTCGATCAACGCGGTGCTCGCGCCCTTCGGCCAGGTTGCCTACGCCGCGCCCAACG
30 CCTTCCTCGACGCCTTCGCCAGCGCCACGAACACGACGAGCGTCTTTTCGCGCTCAGCATCG
GTTGGGACACCTGGCGCGAGGCCGGCATGGCCGTGATGCCGCCCGCGCCCGCGGCGACAGG
CCCCGCTCGAAGGGCTTAGCGACGAGCAGGGCTTGCGCCTGCTCGAAAGCGCCTTGGTTCGGTT
GCGAACC GCGACTCCTCGTCTCCATCAGCGAACTGCGCGCTCGACTAGCCGAGCATCATCGCA
ACGGCGGCATTCCCCGGTTGCTCGGGCCCCGCGCCAACGAGGCGGGTGCAGCTGATTCCGGCG
25 AGGAGGGCGCCACGCAAGACGCGTCGCCGGCCCGTCGCGCCCGTCCCGATCTGGTTCGTGGCCT
TCGCGCCGGCCGGCAACGAGCTGGAGCGCCGGATCGTGGCCATCATCGGCGCTACCTGCGGC
TCGGTCAGGTGGGCGTCGACGACAACTTCAACGATTTGGGCGCCACCTCGCTCGACCTCATCC
40 AGATCGCCCAACGCCTCGGTTCGCGAGTTGGGCCGCGATGTCCCTGTCTGCTCTACCAAC
ACCGCACCGTACGCGGGCTGAGCCGCTTCTTCGGCGGCGCGCTCCAATCCGCGCGGTCCGGCG
30 TCCCCAGGGCGCTGCCGCACCGGCGCGCCACGCGGGGGTTGCCACCCCGCCGCGGCCAC

5

73

10

15

20

25

30

35

40

45

50

55

AACCGTCGCGCCAGCACCTGGAAAAACGCCGTCAATTGAGGAAAAAGGGGGCCTTCCCATC
ATGAGTGAAGTATCCATTGCCCCGGCTTGGACATCGCGGTATCGGCATGGCCTGCCGCTTT
CCCGGTGCCCCGAACCTCGCCGAGTATTGGGCCAACCTGATCGAAGGCCTCGAAACGCTCAGC
TTCTTCAGCGAAGAGGAGCTGCGTGAGGCCGGCTGCGATCCGGTCCAACCTGGCCCAGCACAAAC
5 TACGTGCGCACCAAGGGCCTGCTCCCTGACGCAGACCGTTTCGACGCCGATTTTTTTGGTTAT
TCCCCGCGCGAAGCCCAGGTGATGGACCCCCAGATCCGCGTCTTCCACGAGGTCTGTTGGCAG
GCGCTGGAGCACGCGGGCTACAACCCGCATCGCCACACCGGCACGATCGGCCTGTTCCGCCGGC
GCCGCGCCCAACGTTTTTTGGGAGTTTTCTCTCCTATCGGTCCGATGCCGCCAATTTAGGCAAC
TTCACGCTGGGCCTGCACAACAACAGGACTACCTGAGCTCGCGCATCGCCTACAACCTCAAC
10 CTGACAGGGCCCAGCTACACCCTGTTACCCGCTGCTCGACCTCGATGGTCGCCATCCACCAG
GCCGTCCAGGCGCTGCTCAACGGCGAATGCGACCTGTGCATGGCCGGCTCGGTCTCCATTACG
CTGCCACTGGTTGCCGGCTACACCTACACGCGGGCATGATCGTCTCGCCCGACGGCCATTGC
CGCACCTTCGACGCAGGCGCAATGGCACTGTCTACGGCGACGGGGCCGGCTGGTCGTTCTC
AAGCGGGCCGAGGATGCGTTGGCCGACGGCGACCACATATTTGCGCTCATCAAGGGCTCGGCG
15 CTCAACAACGATGGCAGTCGCAAGACCGGCTACACCGCGCCCAGCGTGCAGGGGCAGGTGGAG
GTGATCCGCGCGGGCATGAACCTGSCGGAGGTGAGCCGGAGGCGATCAGCTACGTGAAACC
CACGGGACGGGCACCACGGTGGGCGATCCGCTGGAGTTCGAGGCGCTAAAGGAGGCCTTCGGA
GGTGGCTGCAAGGCCTTCTGTGGAATTGGGTTCCGGTCAAGCCGAACATCGGCCATCTGGACGTG
ACGTCGGGGATCGCGAGCTTCATCAAGCTGGTCTTGGCGCTGGAGCACCGCATCCTACCGCCC
20 ACGCTCCAATTCCAACCTGCCAACCCGAAGATGGATGGTTCGATAGCCCCCTTCTACATCGTG
GCTGAGCGCGAACCTGGCGCGAAGATCTGCTGCCGCGTCGGGCCGGTGTGAGCGGTTCCGGT
CTGGGTGGCACCAACGTCCACATGATTTTGGAGGAGTTTCAGCGCGAACCGGCGCGAACAGC
GCGCGCACGCGCCACCTGACGGTGTGACGGCGCGGTGCGCGCAAGCCCTGGCGCAGCTGGCG
GCCAACCTCGCCGAACACCTGCGCGAACACCCCGAGTTGGCGCTGGCCGATGTGGCCCATACG
25 CTGCTGCACGGCCGCAAGCCACATCCATTGCGCGCATCCTGGTGGCGACCGATACGACGGCG
GCGATCGACGCCTTGATGAACGACCGCGATCCGCGAACGCGTTTCTTGAAGCGACCGGGCGC
GGCGAGTCGGTGATCCTGTGTTTTACGAAACGCCGCGGAGCCGCGAAGCGCCCGCTACCTC
45 TGGGATCACGAGCCGCTTTATCGCSCGGCGGCGACGTGCTGCTTGGCTGGTGAGGTGCGCGAC
CCGATCTGGAAGGCTGCTTTACTSCCCTGATCGCCGAGCAGGGCGCGGACGCCCTTTTGC
30 CACCAATACGCGCTGGCCGGATGCTGCTGGCCATGGGGTTGACCCCGTCGGCGTTGATCGGC

5

74

10

15

20

25

30

35

40

45

50

55

GTGGGCCAGGGCGAGTGGGTAGCAGCGGCGCTCGCGGAGGTGTTCCCGCCATCGGCCTGCTTG
CGCTGGATTAGGTTTCGGCGAACGGCTCCCGCAGCCGCGCGATCAACGGATTCCGTTTCTCTCC
AATTTCTCTGGAAACTGGATCGTTGGGCGTGAGTTGGCCGACCCGGATTACCCAGAAAGCAG
AAGGGTAAGCGCTGCATGAAGCGCCGTCGGTCCCAACCTCGGTTCAGCTGGTGAGGATGGGGG
5 CGATGGAACCGGCTCGGTTCAGCTCGTCGCGCGCTGCTCTTCCGCGGAAGCGGAGGCGGGACG
GTGATCGCCCCGAGGGCGAGGTTTCATCTCGTCGTGACGAGCCGGGCGCGGTGCGCGCCAG
TACCTGGGGGCGAGCTCGAGGTAGCGGTCCCGCGCCAGTAGGGCATCGCGCGAATGACGTCG
GCCAGGTAGGCCTCCGGGTTCGAGCCCGTGCAGCTTGACGCTCGCCACGAGCGAGAAGAGTTG
GCCGCGGCGGAGGCGTGGTTCGTGCTGCCGAAGAAGAGCCAGGACTTTCTCGCAACCGCAATG
10 GATCGCAGCGCTCGCTCGCTGGCGTTGTTCTCCAGGCGCAGCCGACCGTCGTGAGGAAGCGC
CGCAACGGCTGCTCTTGGTTGAGGGCGTAGCCGAGCGCGGTGGAGACCAGGCCGCGCTCGCGG
GGACGAGCGTGCTCGGCCCTGGCCCCAGGCAAAGAACGCGTCGACCAGAGGGCGGACGACACA
TCGCGACGCACCTTGCGCTGCGCGGGCGGCAGGTCCGCCAGCGCGCGATCGGCGGCAAAGAGG
25 GCGTTGATGCGCCGAGCCCCCTCGACACCGAGCTCGTGCTTGACAGACCGCCGCTCCAGAAAG
TTGGTACGGCAATGCGACCAGCATCCGACTTCGGTCGGGGGCGGACCGCGCTTCTCGTCGGCA
GCAGCGCTCTTGGTGGTGTGCCGCGGAAGAGGGCGTCATAGATGGCGTGAGCGTCAGCTTGA
ATATACCGAGAGAAGCCGCGGAACATCTCGCAGACCGCGCGCTGGTATGCTTGGGCTGGTAC
TCGAAGAAGACGTGATCCTTGTCGCGAGGACGACGAAGAAGTGTCCTTGCGGCACGGCCCCG
GGCTTCTTGTCCTTGCGCTCCTGGATGGGCCAGGCTGGACGGAGACCCCGGTGGCGTCCGTG
20 GACAGGCAGAAGGCGGTCTCGAAGGCCTCTTTGCGCGCGGCTTCGACGATGGCGCCAGGGTC
GCACCGACGTCTTCGGCGTAGCGGCACATCGTGCCGCGATCGAGCGACGCGCCCTGAAGCTCC
AGCTGCTGCTCCAGTCCGATAGAACGGGACGCCGAGCAGGTACTTGCTGGTGAGGATGTGCGCA
ATCATCGACGGCGCGAGGAACGACCGCCGGAACAACCTCCTTCGGAAGCGGCGTCGTGATGAAG
40 ACCGTGCAGGTCTCGCCCTTCGGCGCCGGCGGCGCGCTCGAGCGAGGGCGCGTCGAGCGCT
25 GTGGAGGAAGCGCTCGGCTCGCCGGCCGCTGCCGTGTCCTCCGGGCTGACGCTCGCAGCGGGC
GTCGGCGTCGGGGCTTCTCTCGCAGCAGCTTGAGCGGGGCCGCTTCTCCTCGCCCCAAGT
CTCGCATCCGTGACGGACCGCTCGGCCTTGACACGACGCGTGCGAGCAGATCGGGCGCATT
45 CCGCGCGCTCGTAGCCGAGTCGCGAGGTCTCCTCGACCCCGATGCGCGTCGCGCTCGCATCG
AGCTCGGGGACGAGAGCTCGATGCGGACGACGGGCGAGTCCGACTCGGACAGGTGCGGACGG
30 CCCTTGCCGCCGACCTTCGTTTCGGCCCCCTTGGGGTCGTGCTGCGCTCGCTCGCTGTA

5

75

10

15

20

25

30

35

40

45

50

55

TTGCGCTCGGCGGCGTCGAGTGCCTTCGCGAGGCGCTGGACCTCGAGGAACATCGAGTCGAAC
GCCAGCTGCTCCGCGCTCACCTCGGCGCGCTCCGCCTTGGCCACGAACAGTCGACGTCGCAGA
AGCTGCAGCTGCTCGAGCGCACGGGTGTAGGCGCGCCGAAGCTGCGCGAGCGCATCGCGCGCT
CCCACGAGCTCGCTCTTTGCCGCGGCGAGCTCCGCTTCGAGCTGCGCGATGCGCTGCTGCTCG
5 GCCGAGAGCGTCGGCTTGGCGGCGGCGTCGTGCACGACGCCGCTCTACGTAAGCCGCGCGTAC
15 TTGTCGAGCGAATTCGTGCGGCTCAGTGSACGCGGCGCGGTGCGCGCCTTCGCGGTTTGGACG
TGGGCGCGATCTCGATGCCGTCGAGCAGCGTCTCAGCGTGCGCTCGTCCACCTCGACGTGCG
TGGCGCCCTCGGTCCGGGGGTCCGGAAGTGCGAACGCTCCGCGATCAAGGCGTTTGTAAAACA
GGCAGATTCCACTGCCATCGAAGAAGAGAATCTTGATCGTGGTCCGCCGCTTGCCGACGAACG
20 CGAACAGCGCTCCGCAGCGAGCCTCGTACCCACACGCTCACGGATGAGACCCGAAAGCCGCT
10 CGAAGCCGTAGCGCATGTCCACCGGCTCCAGCGCAGCAACACCTGCACGCCCGCCGGAATCA
TCGCCCCGCTCCGCCGAGGGCACGGACCACCTCCGCCAGCAGCGGGGTGCAACCCCGCGGC
GACGCGCACCCGCGCGCCCGACCTCGACGACGAGCTCCGCAGCGCTGCTCGTCACGGCGGG
25 CGCCTTCGGCACCAGGCGCAGAAAGCGCGGTGGCTCGCCCCGCGACAGCCGGCTCGACCAGCC
15 GTGCAGCGTCGAGGCCGCAATCCGCGGCTCCGAGCGAACTCCTCCGCCGTTTACCACTCTC
GCGCCACGCCCGAACCGCTCGGACCACATCACTTCGGTCGCCTTCGTCCTTGTCATGCACGC
30 CATCATGAACTGGACAGCGCAGCCGGGTGAGACGGCGCTTCGCGCAGCGCTTACGCAGAAGG
CGCGCCGCGCGCCATTGTGCGATGCGGTGCGCGACTTCGCCCGCGATCGGCTGTTGCTGGAAC
TGGGACAACCACTGGACGTAACGGCTGAAGCGAGCCAACGGCTCCAGCTCGCGCGGGGCGACC
20 TGTTCCGGCGCCTACCAAGCGTTGGCCAGCTCTGGATCTGCGGCGCCCTGGCCGAACCGCCGC
35 GACTGTATCCCGACGAACACCGCCGGCGCGTGCCGCTGCCGAGCTACCCCTTCGAGGGAAAGC
GGTTCTGGATCGAGGGCTCGCCGTTTCGAAACCGCGCCCGCCCGCGCCTACCCCAACCCG
CCGATTCCGGGGACATTCTCAAGGGCGACCCGGCGGACTGGTACTATCGGCCGCGTTTTCGAAG
40 CGGCGCCGCTCTTGCCAGCCCGTTTCGAGAGCGAACCCGGCGATTGGCTGGTGTTCGAAGATG
25 AGCTGGGGCTCGGCGCCTGGCTGAGCGAGACCTTGCGCGACAAGGGCGCGGGTTCGCGACAG
TCGTTTCGAGGCACCGAGTTCCGACGCCTGCGCTCACAGCGCTTCAGCTTCGTCCCGATCGAC
GGGACGATTACCGGACCCTGCTGCACGAGTTGAAGGCGCAGGGCATCGCGCCGTTCCACCTGT
45 GCCACCTATGGAGCGTGACCGCCGCACCGGATGCCGAGCAGTTGCTCGACGTCAGCTTTCACA
GCCTGGTCCATTGGCGGCCGCTTTGGGTTTCGGTTGGCTACTTCCACGCCATGAAGTTGAACG
50 TGGTCGCCCAACCGCTATTTCGACCCCGAGTCGCCCCGAGCGCACCGAGCCCGCCAAGAGTCTGT

5

76

10

15

20

25

30

35

40

45

50

55

TGCTCGCGGTGACCAAAGTCCTGCCGCAAGAGGTGCCCAACGTTGGAACCCGCGCCATCAGCG
TGGACCTGGATCGCTCGTTGACGCGGCGGCGCCCGCTGGGCGGCCAGTTTGTGGTTGAAT
GCGGCGCGCCCGTCGAGGAAACGGTGGTGACCTACCATGGCGCAGCCCGATGGCTGCGCCGCT
TCGATCGCGTTGCGGTGAATGGTCTCGGCCCGTTCCACCCCGATCAACCTGCGCCGCTGCTGC
5 GCGAGCGCGGCGTGACCTGATCACCGGCGGCCTGGGCGGCGTGGCTGGCCAGTTGGCGCGCT
ACCTGGCGCGGGCCTGCCGGGCGCGSTTGGTGCTCACCGCGCGCCGGCCCTGCCCGAGCGCG
ACCACTGGGATCGGGAGTCGGCCGTGCTGTATGGGACGACAAGACGCGCCAGCGCATCGAGC
TGGTGCGCGAGCTCGAGCGGCTGGGGGCGGAAGTATTGGTGGTGGCTGCCGATGTGCCCGACG
AAGCGGCCATGGCGCAGGCGATCGAAGCCCTCACTGGCGCGATTGACGCTTTGGACGGCTTGA
10 TCCACGGCGCCGGGATCGTGCGGCTCGCGTCGGGCGCGCACGCCGATCGGGAGTATGACGCGGG
CCATGTGCGAGGAGCAGCTCCGCCCCAAAGATGTTGGGCCTCGACGTCGTGACCGCCTCTCTGC
GCGATCGCCGGTTGGACTTCGCGATTGCCATCTCGTCGCTCGCCCCGATTCTCGGCGGCCCTCG
GCCACGTGCGCCTACGCCCGCCCAACCTCTACATGGACGCGTTGCGGACGCGCGCGCCGCCCG
GCAACGCGCCTTGGATCGCGCTGAACCTGGCCGAGTGGGAATACGAGGGCCCGGCTACCTACG
15 ACGAGCGGGTGGGCCGTTTCGCTCAAGCAGCTCGAGCTACCAACGAGGAGGGTATCCGCGTCT
TCCAGACGGTGTGGCCTTGGCCGCGCGCGGCCCGCTACAGCAGATCATTATTTCCACCGGCG
ACCTCCAGGCCCGCCTCGACAAATGATTACATCAAATCCCTGCATCGCCGACCGGGGCCCG
TCCAGCTCAGTCGCCGACCGCGGCACCCAGGGCGGTTTCGGCTCGGAGCGCGCCGCTTCG
AGGCCGCCCTTCGCTGACGCCTGGTGGGACTTCTTCGGGGTTGAAGAGGTGACCCGAACAAA
20 ACTTCTTCGATCTGGGCGCCAGCTCGCTCGACTTCATCCACCTCGTCAGTCGCTTCAGCAAGG
CCATCGAACAGCATGTACCGCTCGAGGCCCTGCTCGAACACTCCACCTGCACGACCTCGCCG
CCCACCTCGCGGGCGACGCGAACCCGACGCGCAGCGACGAAGCGCGCATTGCGCAACGGCTGC
AAGGCGCCAAGTCCGGCGACATCGCCATCATCGGCATGGCCGGCCGCTTCCCGCTCGCGCCCG
ACCTGGACACCTATTGGCGCAACCTGGTCGGAGGCATCGACGCGGTCAGCTTCTTCAGCGCCG
25 AGGAGTTGCGTGCTGCTGGCGTCAACCGCGGCCGAGATCCACCACACCAACTACGTGCCGGCCA
AGGGCGCTGCGCCGACAGGACTTGTTCGATGCGGCCTTCTTCGAATACACTGCCAGCGACG
CCGAGCTGATGGACCGCAAAATCGCGTGTACACGAGGTGCTGTGGCACGCGCTGGAAGACG
CCTGTTTCGACTTCAACGGCGATCACGGCCAGGTCGGCCTGTTGCGGGGCGCCTCGCCGAACC
40 TGTTGGTGGCAGTTCTGTGGCCAGCTTTCCGAGGCCCGCAAGACGAGGGCATGTTACCACCA
30 CCCTGCTCAACGACAAGGACTCGATCGCGACCCAGATTTATACAAGCTCGGTCTAAAGGGCC

5

77

10

15

20

25

30

35

40

45

50

55

CCGCGGTACCTTGTTACCGGCTCTTCCACCTCGCTGGTAGCCGTTGACGCCGCTGCCGCT
CGATCTGGTCCGGTCAATCGGACATCGCCGTGGCCGGCGCGGTCTCGCTGACTCTCCCCGATA
AGGCCGGCTACATCTACGAAAAGGCTATGCTCTTCTCGGCCGACGGCCATTGCCGGGCTTTCCG
ACGCCAACGCCACCGGCATGGTCTTCCGGCGACGGCGCCGGCGCGATCGTGCTCAAGCCGTTCG
5 ACGGGGCCCTGCGCGACGGCGACCCGATCCATGCGGTGATCAAGGGCTGCGCCACCAACAACG
ACGGCGACCGCAAAGCCGGCTACACGAGCGTCAGCGCCCAAGGCCAGGCCGAGGTGATCCGCT
CGGCCCAGATCCTGGCCGACGTGGCGCCCGAATCCATCAGCTACGTGGAAGCCACGGTACCG
GCACCAAGTTGGGCGACTCGATCGAGATCAAGGCGTTGAAGCAAGCCTTCGCCAGCGACAAGA
ACGGATTTTGGCGCATCGGGTCGGTCAAGACCAACCTCGGTACCTGATGGCGGCGCGGGGA
10 TGGCCGGCCTGATCAAGACGGTTCTCGCGATGAAGCACCGCCAATTGCCGCCATCGCTGCACT
GCGACGAAGTGAACCCGACCTGGATTTGGAGCGCAGTCCGTTCTACATCAACACCCGCTGC
GCGACTGGGTTGCACCGGGCGGGCGCTGCGGGCGCGGTGAGTTCGTTCCGGATCGGCGGAA
CCAACGCTCACGTATCCTGGAGGAGCCCGGACGCGGAGAGCGGCACGCGCATGCCCACT
25 GGAAATTATTGATGCTGTGCGCGGCCAGCGAGGCGGCGCTCGACCGCCAGGCCGATAACCTGG
15 CCGACTACCTGGAGCGCCATCCCGAGCGCCACCTCAGCGACGTGGCCTATTCCCTCCAGACCG
GCCGGCGCGTTCTGGCCTGGCGGCGCACGGTCTTATGCGAGTACCGCGAGGACGCGGTGACCA
GTCTGCGCGAGCGACAGGCCAAGCTCGTCCAGACAAGTCCGCTCCGCTGGGACCACAAGGACG
30 TGGTCTTCATGTTTCCCGGTACAGGCGCGCCAGTACCTCAACATGGGCCGCGACTTATACGTCA
TGGAGCCGGTCTTCCGCGAGGTCACTGACCGCTGCTTCGAGTTGCTGGCCCCCTTTGTGGTCCG
20 AGCATCCGCGCCAGATCCTTTATCCGAGGGCGGGGTGTCGACCCTGCTCCACCGGACTGATT
35 ACACCCAGCCGATCGTGTCTGCTTCGAGTACGCCCTCGCCCATTTGCTGCTCTCCTGGGGAT
TGAAACCGGCCGCGACCATCGGCTACAGCTTCGGCGAGTACGTTTCTGCCTGCCTCGCCGGCG
TCTTCTCCCTGGAAGATGCGATCCGCTCGGTGACCGAGCGCGGTCCGGCTGATGGCGGCTTTGC
40 CCGCGGGCGCCATGCTCAGCGTCCCGTTCCCGAATGCGAGCTGCTGCGGCTGCTGGACGGCT
25 TCCACGCCCAATCGGCGGCCCATCTCGCGCTGGCCGTGACAAATGGCGCCTCCTGCATTGTGG
CCGGCGAGCAGGCCGCCATCTCGGCTTCGAATCGATGCTTCGCAAGAAGCGTCTGTTGACCA
TGCCGTCGCGGTGAGCACGCGCTTCATTGCGAGGTGATGACCGGCGCGACCGACGCCCTGC
45 GCAGTATCCTGCGGAAGATCCCCCTCTCCGCGCCGACAATTCCTTCAATTCCTGCGTCACCG
GCACCTGGATCACTGCACAGCAGGCTACGGATCGCGAGTATTGGGTGAACCACATGTGCGGGA
30 CGGTCCGGTTCGCGGCGGGTCTGACCGAGCTGGGTCAAACCGCGAGGCGGTGTTCTGGAAG

5

78

10

15

20

25

30

35

40

45

50

55

TAGGTCCGGGCCGCGACTTGACGTTGCTGGCCACCGCATCCTGGCCGACAGCGCGCCGTGT
TCGAGCTGGTCAAGGCGCCCGACGGCGGCGACGACGATGGGTTCTCCTGCTGGATCGATTGG
CCAAGCTCTGGAGGCTGGGGATTTGATTGACTGGGCCGGCTTCTACGCGGATGAGCGGCGGC
GGAAACTCTCGCTGCCGGGATATCCGTTTCGAGCGGCGGCGCTTCTGGATCGAGGGCAACCCGC
5 TGGAGATCGCCCGGCGAGGCCCAATGTCCAGGGGCGCTGGTCAAGGCGTCGGACATCGGCG
CTTGGTTCTACGTGCCGCAATGGCGGCGGTGGTGCTCGCCGAGCCGGGTACAACGGCGGCGG
GC3CCGCCGTACGGCGGAGCAGGCACGCGTCGTGACCGAGCTACGGGCGGGATGCGCGTCGG
CCGGCTTGGGCAGCGGGGCTGCGGACTGAATGGCGGTGCCCCGTCCGAGCGTCCGAAAGAAA
GTGTAGCGCCAGCCGGGTGACCA3CGCAGCGGCGCAGACCGGCGCGGACTGCCCGACACCGA
10 CTGGGAGCCAGCGGCTGTGCCAAAGGACGGGCGGAGCCGCGGCGGACCTGGCTTATTTTCG
CCGACGCCGCGGATTGGCCGAATCTTTCCGAAGCGGGTTCAGGCCCCGCGCGAGAAGCTTT
ACCTGGTGGCTTCCGGCTCGCGCTTCGAGCGCTGGCCGAGACCCGCTTCCGCTCGATCCCG
GGGCCAAGTCCGATCACCGCTGCTTTTCAAGGCGCTCGACGAGGCCGACATCCTGCCGACCC
ACCTCCTCGACTTCGCTCGCTTGACTGCGCGGGGCGGACCGGACCCCATGGACCAGGCCG
15 GCTTCTTCGGGCTGTTGCACCTGGTCCAGGCGATGGCAGAGGCCGGCTACAGCCATCCCATTC
GGCTGCTGATCGTCAGTTGCGGCGTCTACGATGTACCGGTGCCGAACCGCTGCAGCCGCGCG
GGGCCACGATGATCGGACCGGCTCTGTGCATCCCGCAACAGTATCCGCACCTCGAAACGAGCC
ATGTGGATTGGGCGTGGTCCATGCCGACGAGCTCCACGCCGCGCGCCAGCTCGACAGCCTAC
TTGCCGAATGCCTAAGTGCAACGGCCGAGCGCCAATTGGCGCTGCGCGGCCGACACCGCTGGC
20 TGCTGGACTACGAGCCAGTCCGCTTGCCGCGGCTCGACCCGGGCGCTTGGCCCTGGCGCCAGC
GCGGGTCTACTTGATCACCGGCGGTTTGGGCGGGATCGGCCGATCCTGGCCGAACACCTGG
CCCGCACGACCTCGGCTCGCCTGGTCTTAATCGGCCGCGAAACCCTGCCCGACCGCGACGACT
GGGACGCTTGGCTGAACCGCCCGCAACCGGTGACGCGCACCCACGAACGGCTGCTGCACAAGA
TCCGCGCGATTTCGCGATCTGGAAGCGCTAGGCGCCGAAGTCTGCTCCTCGCCGCCGACGTCG
25 CCAACGAAGCCGCCATGCGCGAGGCCTACGATCGCGCCGAATCCCACTTCGGCACATCCACG
GGGTGATTACGGCGCCGGCCTGATGGACGCGCAAAGCTTCTCACTGATCGACGCCCTCGACC
ACGACCTCTGCGCCCGCAGTTTGAAGCAAAAATCCGCGGCGTCTGCGTGCTCGACCGGTTT
40 TGGCCGACCGCACGCTCGACTTCTGTCTGCTGATGTCTTCCATCTCCACCGTGCTCGGCGGCC
TGGGCTATTTGGTTACGCGCGGCCAACGCCTTCTCGACGCCTTCGCCAGGCGCGCAGCC
30 GCGACGCCGCTTTCCCTGGCTTAGCGTGGCCTGGAGCCATTGGAAGTACTGGACCGAGCGCA

5

79

10

15

20

25

30

35

40

45

50

55

AGATGGACAACGAGGTCGGCGCCGTCATCGACAGCCTCTCGATGGAACCCGCCGAGGGCTTCG
AAGCCGTCACCCGCGTCTTGGCTTGGGGCAAGGCGCCCCACATCGCCAACTCGCCCCGGTGACC
TCGGTCGCCGCCGGGATCAATGGGTCAAACCTGGCCAGCCTGAAATCGGCGCACTCCAGCGAGC
CCGAGCCGGCTAGGCATGGACGTCCGGCGCTCTCCAGCGAATGGGTGCGCGCCGCGCAACGTGG
5 TCGAAGAGAAGCTGGTCGCCATTTTCGAGCAGGTGTTGCGCACTGCGGCACTGGGCATCGAGG
15 ACAACTTCTTTGAGTTGCCCGGCGACTCGCTCAAGGCGGTGATGACCGCGGCCGATTCAAA
AGGAGCTGAACGTGGAAGTGCCGCTGCCGACCTTCTTCCAGATGCCACGGTCGCTGGCCTGG
CCGAGTTCTGTGACGCAAGCCAAGCGCAGCGGCCGGGAGACGATTGCGCGCACCGCGCCGCGCC
CACATTACCCGCTCTCGECTGCCAGGGCCGCCATTACCTGCACTACCGCATGGACCCGCGTT
20 GTACCGCATACAACGATCCCTTCGCCAACCTGATCGAGGGTCCGCTGGACGTGGATCGCGTGG
AGCGCATCCTGCACACCCCTCATCTACGCCACGACTGCTTCCGCACCTCGTTCCACTTCCGCG
AGGGCGAGCCGGTCCAGGTGATTACGATCGGGTGGACTTCAACCTGGCGCGGATTACCTGCG
CGCCCGAGGATTGCCCCGAACGGATGCGCGATTTCATCCGCTCCTTCGATCTGGAGCGACCGC
25 CCGCCATGCGCGCCGGCCTCTTCTGTCACGGGGCCCGAGCGCCACGTGCTGCTAATCGATTTTC
15 ACCACATTATACCGATGGCGTGTGCTTCGAGAACTTCGTCGGCGAGTTTCGCGGCGCTCTACC
GCGGCGAGATCCTGCCCCGAGCTGGAACTCGAGTACAAGGATTTCGCGGTGTGGCAGCATGAGA
30 ACCGGGGCCGCGCGCCAACAGCGACCAGGCCCGCTACTGGACCGAGCAGTTGGCCAATGCGC
CCGGGCCGATCGAGCTAACCACCGATTTCCCCCGTCCAGTCGACGCGAGCTTCCGCGCGGACC
GCGTGCGGACCGTGCTTGATGCGGAGCTCGTTGCTCGACTCAAAGAGCACGCGGCGCGCCTCG
20 GCATCACCCCTCTATAGCTGTGCTGGGCGGATTCTCGTTATTGCAGCACAAAGCTCTCCGACT
35 CGCACGACATCGTCATCGGTTGCCCCGTGCGGGCCGCGACCCGGAGCGAACTCCAGGATCTGC
TGGGCGCGTTTCGTCAACACCCCTGCCGATGCGCCACCGCATCGACCCGACCCATACCGCACGGG
TCTTCTTGAGCAGGTCCACCAGACAACCTTGGCGGCCCTCAGCTACCAGGAGCACCCCTTTTG
40 ACGAAATGGTGGCGACGCTCGGGTTGCGCGCGGATCCGGCTCGCAACCCGATCTTCGACACGA
25 TGTTCTTGCTGCAGAACATGGCCATGGGTGCAACCACCTTCCCGGTCTGCGGCTCTCGCCTC
ACGACACTTTTACCGCAAGGCATTGTGCGACCTGATGCTACAGGCGACCGAGTATGACTGCC
ACCTGGAGCTGGTGCTCGAGTTGCCACCGACCTGTTCCGGCTGGAACCGCGCAAGTCTTGC
45 TCGACCGCTACCGCCAAGTCTTGAGTGGCTGTTGGCGTACCCCATGAATCGATAGACGATT
TGACGCTCGCCGGCCACTTTCGCGAAGTCGAAGTGACGATGTCGGACGAGGGCGACTTTGATT
30 TCTCAGATTTTGAACCCCGCAACGTGAGAAACCTATGGCGCGCTGAGCCGCACAGATCTCCA

5

80

10

15

20

25

30

35

40

45

50

55

ACTCGCCATTCAACCAGCGCACCGTGGAGCGCGAATATTGGCGCGCTCTGTTCGAGCGCCATCC
GCAACGGTCCAGTTTGGCGGGGGTCTCACCGCCCCGATCGGCGACGAGTCGACCCGCGAGAC
CTTGTCATTCTGTCCTCGACGAAGATCCCCTTCGGCTGAGTAATCGTTCCCGCAACGCCTGCT
CACGGTGTGGCGGCTGGCCTCGCGGCTTCTCTCCACCGCTGCGACGGCGCTGAGCGCTTCAC
5 CCTGGGGTTGGCCCTACCGCGCCAAGCCGATGACCATCACCCGATCCTCAACAGCTTGATCGC
15 GCTGGGGTTCGCGGTCGACTCGAGTACGACCTTCCGCGATCTGCTCTATGCGCTTCGATCCGA
ATACCACGAGGCGATGCGCCACGCCAACTTTCCGCTGGCGACCTGGTGGCGGGCCCTACCCGG
CGGAACGGCGCCGTTTCGACGTGCGCCTCAGCCTGGACCCCTTCACAGACGGCGATTCTGCTGGA
AGACCACGCGATCGGCGCTTGTTCGGTTTCGATTGAGGGTGAGCGCCTCACCTGCCGATT
20 GCGATTGACCCCTGCGCGCTATGACCTGCCGCGATCGAAAACCTCGCCGATCGTTTCCGCCG
CTTCTCTACGCGCCTGTGCCGGGACGCCTCCACCGTCATCCAGGCGCTGGACCTTTCGCTGCC
AAGCGATGAATCGGTGTGGCGGTCCTGAAGGCGTGGCGCGGGCTATTGCAAGACCTGAC
25 GCTAGACCGCGCTTCCGCCGCCAGGCCGCGCAAACGCCCGATCAGCCGGCGATCACGTTGAA
CGGGGACGTCCAGAGCTACGCCGAGGTGACCGCCGAGCGACGCGCTGGCCCGCCACCTCCG
15 TCGCCACGGCGTTCGGTCCGGAACGATTGTGGCCGTCAACGCCCGCGCGGGCCTAATCAGCT
GACGGCCCTGCTCGCGGTCCATAAGSCCGGCGCGCTACCTGCCGATCGATGCCGAGGAGCC
30 GGCTGCCCGCCAGCAATTCAAGGTGCGCGACAGCGGGCGCGGTGGCACTGGAGCCGTCGCC
GGACCAGGCGCTGACCGTCACCGAATCGCCGCGGTCTTCTTGACGATGCCTCGCTCTTCGC
TGACGGCGGGCTCGATGTGCCGCGCGCGCCGACTCGCTCAATCCGGCCTATGTGATGTACAC
20 GTCCGGCTCGACCGGACAGCCCAAGSGTGTGGTGGTTCCCCACCGCGGCGTGGTCAATCGTTT
35 GAATTGGGGGCGAGTCCCGTTTCCCGCTGGACGAACCGGACCGAATCCTCCAAAAGACGCGCT
GCTGTTGACGTGTGGTCTACGAGCTGTTCTGGGGCGCATGGAGCGGGGCCACCTGACAT
CCTCGAGCCCGCGCCGAGCGGACCCCGACGAGTGGCCAGGGCCCTGGCCGAGCGCGCCAT
40 TACCGTATGCCATTTCTGTGCTTCGATGCTGCTCGTCTACTTGAAGTCATGCGGCGGCACCA
25 TCGCGCCCGCGTCCCGGACCGGCTCCGTTACGTCTTCGTCACTGGCGAGGCCCTCGAACCAGGA
CCACCTCGCCGGGCTCCAGCAGATTGGTTCGGCGCCTCGGCCGACGATTCCCTCGTTAATCT
GTATGGACCAACCGAGGCCTCGATCGAAGTCTCTGCTTCGCTGTCCCGCCGACCATGTGCC
45 GCGCCGATCCCCATCGGGCAGCCGATCGACAACGTGCACTGCACGTTCTCGACCGGCGCGG
CCGTCGCCAGCCGCCCTATCTTCTGCGAGCTGTTCTGGCCGGGACTGCCTGGCGCGCGG
30 CTACCTCAACCGTCCCGACCTGACCGCGCTCCACTTCGTGCCAATCCCTTCGGCAACGGCGA

5

81

10

15

20

25

30

35

40

45

50

55

GCGCATGTACCACAGCGGCGACTTGGCGCTCGTGC GCGGCGACGGCCAAGTGGCGTTTCTCGG
CCGCCGTGACCACCAAATCAAAATCCGTGGTCAACGGGTCGAACTGGGCGAAATCGAGAGTCA
TTTGCGCGGGCTCGAAGGCATCGCCGCCGCCGTCGTCCAGGCCGAGTCGCAGCACCATGAAAC
CCTGCTGCACGCCTACGTCTGCACCAACGACGCGGGCCTCAATGCGGCCCGGCTGCGCGCCGC
5 CCTCGCTCAACATCTGCCCAGTACATGATTCCCCAGCGCTTCTCGCGGCTGGCCGAGTTGCC
GCTGCTGGCGGCAGGCAAGATCGAACGCGCCGCCCTCGCGCAACGTGCAACGCCGCTCGCCAG
CGGCGCGCCCTTCGTGGAACCCAGCGGGCCCAACCAGCAGCGTATCGCAGAACTGTGGCGCCA
GGTCTTAGCGGTGCGCGAAGTCGGCGCCGAGGATCCCTTCTTCAGCATCGGCGGCAACTCGCT
CAATGTGCTCAAGCTCAGCGCGCGCTGAGCGACGCCTTCGCGCGTGACATTCCCATGCCGGC
10 CCTGTTCCAATACGACACCATCGCCGCCAGGCCTCCTGGCTCGACGGGCAGGTTGACGAACG
GGCCCAATCCGCCGCGCTCGACCGGCAGGCCGCCGAGGCGGCGCTGACCCCTCAAGAGACCGT
GGCCATTTTGTAGGGATTTCATGACGAACCATGACCATCACGAGGAGAGCAGCGGCCTGGAGA
TCGCCGTCATCAGCATGGCCTGCCGATTCCCGGGTGCTGCCGATTGCGACGCATTCTGGGAAA
ACCTGATCAACGGGACCTCCTCGATCACCATTTCAGCGACGACGAGCTGATCGCGGCCGGCG
15 TTGACGCGCGGACCTGACGCCGAGTACGTGCGCGCGGCCGCCAGATCGATGACGCCGAAC
GTTTCGACGCGGCCCTTCTTTGGGTACTCCAGCGTGAGGCCGAGCTGATGGACCCCCAGTTCC
GCCTGCTCCATGAATGCGCCTGGTCTGTCTGGAACAGGCCGGCATCGATCCGCGCGTCGAAG
CCGCGCCGATCGGGCTGTATGCCGGCGCAGCCGACACCTACTGGAACGCGCTCTCGTCGC
TCGACCGGGGCTCGGCCGAATCGGAGCAATTGCGCGCCGAACAACCTTGCAACCGCGATTTTC
20 TGTGCACGCTGGTCGCCGCCGCGCTCAACCTGAAAGGCCCGCGGTGGTGGTTCAAAGCGCCT
GTTTCGACCTCGCTGTGGCGGTCCACTCGGCCTGTCTGTCGCTCCTGACCGGCGAATGCCGAG
TGGCCTTGGCCGGTGGGGTGGCGCTGCGCTTCCACGCCCCGAGCGGTTATCGCTACGAACCTG
GCATGATCTTCTCGCCCGACGGGGTGTGCCGGCCGTTTCGACGCGGGCGCTAACGGGACGGTGC
40 CCGGCGAAGGCGCGGGGCTGGTAGCGTTGAAGACGCTGAAACGTGCCCTCCAGGACGGCGACA
25 CGATCCACGCCGTGATTGCGCGACCGCGGCAACAACGATGGTGCCCGCAAGACCGGGTTCA
CCGCGCCACGCGCCACGGCCAAGCCGAAGTCATTGCGACGGCGCTGCGCCTGGCCCGGGTGC
CGGCGCAATCGATCGACTACGTGAGGCCACGGAACCGGCACGCCGCTAGGCGACCCGATCG
AGGTAGCCGGCTTGGTGGAGGCCCTTCGCCAGCGAGAAGCGCGGCTATTGCCGGCTGGGCTCGG
45 TCAAATCCAACCTTGGTCATCTGGAACACTGCTGCCGGCATCGCCGGCCTGATCAAGACCGTGC
30 TGGCGCTCGAGCACGCGCACATCCCAAGTCTGCCACGTGCCACGCCCAACCCCGCGCGC

5

82

10

15

20

25

30

35

40

45

50

55

GCCTACACAAGACGCCTTTCCGCATTGCCGCCGACGGGATGGCCTGGCCGCGGCGTATGGCGA
CGCCGCGGCGGGCGGCGGTGAGTTCTGTTCCGCATCGGCGGCACCAACGTCCACGCGATTTTGG
AGGAGGCGCCGCCCGCGCGCCCGAGCTGGCGGACGGGCGCAGTCAGGTGTTCTGTTCTCTCCG
CCAAGGACGAGGCGGCGCTGGACCGTGCCCTTGCCAACTATGGTGCGGCCTTGEAGAAGCGCG
5 GCGACCTCGCGGCGGGCGGCTGGCCTGGACGCTCCAAAACGGCCGGGCCGATTTCGAATGGC
15 GAGCCAGCGCGGTGGCATCCGACCTCGACGAATTGGCGGGCGCATTGCGCGGCGAGCGGCCCCG
GCGCCGTCAAGAAAAACCGAATGGCGCGCGAGGATAAGCCGGTGGCGTTCTTATGTTCCGGGGC
AGGGGAGCCAGTACCGTGGCATGGGCCACGACCTGTACCGCGAAGAGCCGCGTTTCCGGCACC
ACCTCGACGCTGCGCTCGCCATCCTCGCCGAACACAAGCCCGAGATCGACTGGCTGGCGTTGC
20 10 TGGGCTACCGCGACGAGGACGAGCCAACCGACAGATCGGGACGTCCTCGCAGGGCCCCGAGCC
GGTCAGCCGCATCGAACCCAGCGGAGCTCCTCGACAGCACCGAATTGCCCCAACCTTTGCTTT
TCTCCATGTCCTACGCGCTCGGTGCGCTGTGGCTCGACTGGGGCGTCCGACCCACGGCGATGA
25 TCGGGCACAGCCTGGGCGAGTACAGTGTGTCATGTATTGCAGATTTCTATGCACTCGATCAGG
TGCTGCCCTTCATTCTGACCCGCGGTGAGTCATGGCGCAATTGCGGCGCGGCTCGATGTTGG
15 CCGTCAGCGGTGACAGCGTTCTGATGCGCGAGCTGATCGCCGATGCGCTCGATTTGGCGGCGA
TCAACGGCGCTGACCAATTTGTCTGGAGCGGGCCGAGCGAGGCTGTCCAAGCCGCGGGGGTCC
30 GACTGCGCGGCGCCGGCCTGCGTGCCACCGAGCTGAACACCTCACACGCGTTCCATTAGCCA
TGATGGATCCCATTCTGGAGGAGCTAACGGTTGCCGGTTGCGGACTTCAGGTGCGTGTGCGGA
CGATTCCGGTCTGTTTCATGCGTTACCGGAACCTGGTTGACGGCGAAGCAGCTGGCCGATCCGC
20 GCTACCACGCGCGTCACGCGCGGAACCGGTGCGGTTGCGGCGGGCCTAGCGACGCTGACAG
35 GGGAGGAGCCGCCGCTGATGCTCGAAGTGGGGCCGGGCTCGACCCTGGCGGCTTTGGCCCCGG
AGCATTGCAATGCCCCGCTCCCGGTGCTCACCAGCCTGCGCCACGCTCGCCAGGCGACGCCCCG
ATCGCCAATACCTGCTCGAAACGCTCGGCTGCCCTTTGGCGACACGGGGTTTCCGTCGATTGGG
40 GGGCCCATGCCGACGTTTCGCGACGCTTGTTTCGCTGCCCGGCTATCCCTTTCCGGCGCGG
25 TCGCGCGCTTAGCCGGCGACCCCTCCGCTGCTGGCCGGAGCCCGCGCGCTCGCCGCCCCGT
CGGGAACGCGCCAACCTCAGCGCGGACGCGCGGACCTCCCGAACACTCCGGAGCCGACATCCG
GCGCCGTGTCGGCGATCAAAGCGCCAATCGCCGCCGCGATCCCGGCTCTATCGCCTCTCCT
45 GGCGCCAGGCCGGAACGGCGCCGCTCGGTCCGCCCGATCTCGGTCCGCCCCGCGACTGGATCG
TCTTCGCTCTGATTCTCACCTGCTCCAGGCGCTCAGGGCCAATCTCGGGACGCGCGCTCAGC
30 GGGTGACGCTGGTGACGCCGGGCCAGGAGTACGCAGCCGAGCCGTCCGGGTTTCGGCTGCGGC

5

83

10

5

15

20

10

25

15

30

35

40

45

50

55

CGSACCAGATCGACGATTACCGCGCCCTGTGGGCGGACTTGGCGCAAACCGGTATTGTGCCAC
GATACATCGCGTTTCCTCGCCCCGTTTCATGTACCGGGCGCGCATGGCGGGCGATGCCTCGACCC
TGGACGAAGTGGCGGAGGGCGGCTTCCTGCCCCCTGACCCGCTTGATCCAGACTCGCCCCGCCAG
GCGGACCGAGCGGACTTCTAAGCCTCAGCATCGTCACCCCGCGCCCTGGCGCTGGGCGACG
AAGCGACGCGCCCGGAATGGGCAATCCTGCACGGGATGGTCGCCGGCTTAAGCCGCGATTATC
CCGAATGGCGCTTCGTCTCGATCGACGGCGGCGACCCATCCCCGCATCGGTGCGAAGGTCTGG
CCCGCTTGATCGCGCTTCATGCGGTTCGACGAGGCTGGCCCGACCCGCTTGGCGCTGCGCGGCC
TTCACGCTTGGGTTCCACAGTGGCAGCAGTTCAGCCGGCCACCATCCCTGGGGCGGGTATGT
GGCGCGAGGGTGGTGTGTACATGATAACGGGCGGATTGGCGGGATCGGTCTGGCGCTGGCCC
GCGCCCTGGCTCGAGAAGCTCGCGCCAAGCTGATCCTGGTCGGCCGAAACCTGCCACCGCGC
CGATCGATCTCGAGGCTTGGGACGCGCCGCGTGTATTCTACCGCGGACGTGCGCGACGAAG
AGGCCATGCGCCGCGTCTTCGATGCGCGCACGCGCGTTCGGCGCCATCGACGGCATTCTTC
ACGCGGCGCGGTGTCCCCGCTGGCAGCCTGTTTCGCCAACCAATCGGACGCGGCCTTCGAAGACG
TGCTGCACGCCAAGGTTTCGCGGTACCCCTCGTGCTGCAAGGCCTGAGGGCAATCGATGCGCCGC
TGTTGCTGATGTCCTCGCTGGACGCTGGCTTCCCGGTCCCGGTCAGACCGCTATGCGCGCG
CCAACGCCTTCCTCGACGCCTTCGCCAGTCTGCGCCGGCGAGAGGGAGAGCCGGTGTACAGCG
TTGGCTGGGACAGTTGGTGCAGGTGGGCATGGCTGCTCGGGTCGCTGCCCCGATCGGCCGACG
AACGCGGCGCCTGGCGCGCGAGGGGATCAGCCCTCGCCAGGTTGGCAGGCTTTGAGCCGGG
CGCTCGCCCTCGACCCCCCCCACCTGATGATCTCGCGACCGACCTGACCTCGCGCTGGCACA
GTCGATCCAGCCCTACGCCGGTCGCTTCGAGCGAACCAGAGGTGGCGCTGCCGCGCTGGACCG
CATCCGCTGCCAAGCCGTTCATCGACGCTGTTGGTGGAGCACTTCGCCACCGCCGCGTGC
CTCCCGATGGCAACTTTTTCGAGCTCGGCGCCAGTTTCCTTCGACATCGTCCAGCTCAGCGCTC
GACTTCAACAACAGTTTCGGCCGAGATGTAGCCACACCGTGCTCTACAGTCATCCACCGTCG
CCTTGCTGGCCGGCTACTTCGCCAATGACCCGACGCCGTCCGGTGCTGCTGCCGACGAACGCG
ACGAAGCGGTGCGTCGCGGCCGCGACCTCTTGAAGAGCCCGCGGAGGAGTATGACCGTGGA
GCACGAAACCGGATTTCGAAATCGCCGTTCATCGGGCTGGCTTGCCGCGTTCGCGCGCTGCCGA
CGTGCCGCTTCTGGCGCAACCTGGTCGAGGCGCAAGGAGAGCGTGCGCTTCTTCGAGGACCA
CGAGCTGCGGGCCGCGCGGTGCCCGAGGAGATCTTGCGCCTGCCCAACTACGTGAAGGCCAA
GCCACTGCTCGCTGATGGCGAAGCTTTCGACGCGGACTTCTTCGGGTTCATCCGCGCGAGGC
CGCCTACCTGGACCCGCAAGTTTCGCTCCTGCACGAATGTTTGTGGACCGCGCTGGAGGATGC

5

84

10

15

20

25

30

35

40

45

50

55

CGGCTACGATCCCGCGCAGTACGCCCTACCCGATCGGGTTGTTTCGCGGGCGTCTCCAGCAATCT
CTCGTTCCTGTTTCGACCGCATCGATCCGCGCGACTCCCCCTGCAGAAGCGCTATGTGGCCGA
GCTGAACGCGGCCTCCTTCGCCACCCAGATCGCCTACCGGCTCGATCTGAAGGGGCCGCCAT
TTCGATTCAAACCGCCTGTTTCGACGTCCTGGTGGCGATTACCTGGCGGCGCAAAGCCTGAT
5 CGGCGGCGAGTGCCACATGGCCTTGCGCGGCGAGCGACCTTGGAGGTCCCCAAAAAGCCCCG
CTATCTCTACCGGAAGGCTACATCAACTCGCCGGACGGCCACTGCCGGGCCTTCGACGCCGA
CGCGGCCGGCACCATCTTCGGCGACGCGCTCGGCATCGTCTGCTCAAACGCTACCGCGACGC
CCTACGCGACGGCGATCACGTGTACGACGATCAAAGGCTCGGCGATCAACAGTGACGGCCA
TCGCAAGGTGTCTACACGGCGCCGCGCAAGAGCGGTCAAGTGGCGGTGATCCGCGCTGCGCT
10 GCGCGCGGCCAGGTAGAGCCGCAACCATTCGCTTCGTGAGGCCCCACGGGACCGGCACACT
CGCCCGCGATCCGATCGAGGTAGAGCGCTTGACGAGGTCTTTGCCGAAGCGGTGCGGTAC
CTGCGCCCTGGGTTCGGTGAAGACCAACATCGGCCACTTGGATGTGGCGGCGGCGTGGCCGG
TTTCATCAAGGCGGTCTTGCGCTCGAGCGGCGCTCCTCCCGCCAGCCTTCACTTCGTCCG
25 GCCCAACCCGGCCATCGATTTCAACGGGCCCTTCTACGTTTGTGCGCAAATCGAGCGGTTGAC
GGAGAACGGGCGGTTGCGGGCCGGGTGAGTTCCTTTGGCATTGGCGGCACCAATGCCACGT
GATTCTGGAGGAAGCGCCGGCGCCGAGGCGAGACTGCCGGCCGGGAGCCCGCAGGCGCGAG
30 TCCGTTCTGTTCCTCGTATCGGCCAAGACGCCGATGCGCTGGCAGGCCGTTGCCACGACCT
TGCCGACCACCTGCGGGCGCACCCGAGCTCCTCCTGGCCGATGTGGCCCTCACTCTGCAGAT
GGGGCGGGCGTCTTCGCTACCGCCATGTGGTCCAGGCTGCGACGGCGGAGGAGCTGATTCCG
20 CGGTCTGGGAGCGTTCCGACAGGATCCATCCGCAAGAGGCGGAATCGAGTACAATGGGTGTT
GGCAGGCGAGGCGATGTGCTTGACGCGGTTTGCGGCTGTACGCCGATTGGCCGGTCTATCG
GGAGCGGTCGACGTCTGTCTGGCGATCGTCCGCAAGCTGCGCCAAATCGACGGCCGGTCATT
CCTACATGAGTGGATCGAGCGACCGCGCGAGGTTCTGCGCAATGGTTCGACGGCGCTGGCGTT
40 CATGTTCCACTGCGCGCTGGCGCAAGCCCTGAGCCAGGCCGCTGCACCCGACGCGCATGTG
25 GAGCCGTGGGCTGGGCGACAGGTCGGCGTGGTTTTGGCCGAATCCCTGTGTTGGAACAAGC
GCTGGCGCTGGTGTGTGCCAGACACCGGTTCCCGCGATGCCACACCTCAGCGCAACGCTT
GGTTCGACACTGGAAGGCTGCCGTTTTCGTCCACCACGATTTTTGATTTCGGCAGACAGCTC
45 GGGTCGACCCCTGGACCTCGCCGAATTCGCTCATGTGATTTTTGGTGGGTGGCCAAAGCGC
CTCGCCCAATGAGGCGGAGCTGCGCTCATGGAGCGACGCCGCGCCGAGCTGGTGACCTTGGC
30 GATCGGCCCATCCTTTCTCGAGGCCGCTCCGGGACGGTGGGTCTGGCGATCGACCCCAAGCG

5

85

10

15

20

25

30

35

40

45

50

55

ACCGATGACCTGTGTTTCAGCGCACGGTGGCCGCGTTGTGGGAATGGGATGTGACGTGCGCTG
GGCTGCGTTACCTCGTCGACCGGGCGTCGGGTTCCCCTGCCTACCTATCCCTTCGTGCGGGT
AATTCACGATCGGCGACCCCTTCGCGGAGCAGGCGCGGAGGATGACTTGATTGCGGCGAG
CGCTTCGCGCTCGGCCGGATCGCCGCCCGAGCCGTCGGCAAACTCGGCAGCGGAACGCCACG
5 CGCCAGTCAAGCATCGCTCGGCAACCACACCGGCTCCGTCTCATACGTGGCCAGCGTGGC
CGTGGCCACCATTTCTCGAAACCGTCCGTGCCTATTTTCGGGTTTCGCCGCCGTGCGTTCCACCGA
CGCCTTCTTCGAATTGGGCGCGTCTCGCTGGATTGTGTC AACCTGGGCCAGCTCCTTTCCGA
TCGTCTCGGCCGCGAGGTTCCGACCCTGCTCTTACGACCACCAACACCGGACAGTTGGC
GCTGCCCCTGACATCCGCGCGCTCAGCGCAGAGGCGCCGCCCTTAAGGGGCGGTTCATCGCGC
10 ATCGACTTCGGGCACAGCGCGAGCTCGGCCGCTCCACCGACCGACGTTCCCGGGGGACGC
TCACTCGCAGCCAGCTTCGTTTCGCGAGCAGGACATCGCCATCATCGGGATGGCCTTCGGGG
ACCGGGCGCCGACGACCTGGACGCGTTCTGGAACAACCTGGTCGAAGGGGTGAGTCGATCAC
CTTCTTCAGCGAGGACGAGCTGCTGGCGCGGGCGTCCCCCGCAACATCTGGCCTCGACGCG
25 CTACGTGCGGGCCAAGGGGGAAGTACTGGGATGATGGATTTCGAACCGGAATTTTTCGGTTA
15 TTCGGCGCGCGAGGCGCGGTTCATGGACCCGAGTTCCGCGTGTTCACGAATGCTCCTGGCA
CGCACTGGAGCACGGCGGCTACGATCCGACCCGATGCGCGGCATCGATTGGCGTCTACGCCGG
CGTGACCAACCACCTGCCTTGGCTGATGCGAACTTTGCCGCACCTGACCGAGGAGGCAATT
30 CGGCGCGCTGCTCCTCACCGACCGGAGTTTTTCGCACCGCTGCTCTCCTACAAGGTGCGCCT
GCGCGGACCCGCTATTTTCGCTGCAAACCGCCTGTTTCGACGTCGTTGGTGGCGATCGGCACGGC
20 CTGTGCGGAATTGCGCGCGGGTGCCTGTGATGGCCCTAGCGGGCGGCGTGACGGCCAGCAT
35 CGA3CGCTGCGGCTACTTCCACCAAGAAGGCTACATCCTCTCGCCTGACGGCCACACGCGCAG
CTTCGACGCGGCGGCGCCGCGGCACGGTCTTCGGCGACGGAGTCGGCATGGTGTGCTGAAGCC
GCTGGCCCCAAGCCTTGGCCGACGGCGACACGATCCACGCGGTGATCAAGGGAATCGGCATCAA
40 CAACGACGGCGCGCGCAAGGTCGGCTTCACCGCACCTAGCCGGGCGGTCAGACCGAGGCGAT
25 TCGGCGCGCTGCGCGACCGGGGTGGCGTCGAACCGCGTCAGCTACGTGGAGGCGCATGG
AACC3CGACCAAGTGGGCGACCCGATCGAGGTCGAGGCCTTGACCCAAGCCTTTCGCGCCGA
AGCC3ACGGTCCGCTTCGCCCCGCTCCTGCCTACTCGGCTCGGTGAAGTCCAACGTGGGCCA
45 CCTGAACGCGCGGCGCGGCGTGGCTGGTCTGGTAAAAACCGTCTGGCGCTCCAACACCGCCG
CCTGCGGACCGCCTGTTCTACCAAGTCGCGCAATCCACACATCGACTTTGCGGCGAGTCCGTT
30 CC3GTTGAACGGCCAGACTTCGATTGGGTGCGCCAGAGGGGACGCGGTGCTGGCGGGAGT

5

86

10

15

20

25

30

35

40

45

50

55

GAGTTCGTTCCGGTATCGGGGGAACCAACGCCCACCTGATCGTCGAGGAGGCGCCGAAAGCGCT
ACCGACGACAGCGGCACCTCTGTGACGGAGCCGAATGACCTCGACGCGGGCGACGCCGACGG
CCTAGTGCTGCCGATCTCGGCCCGCACGCCGACCGCCCTGGCGCACATCGCGACCAACCTCGC
CAATCACCTGGAACGACATCCGACCATCGCCCTGGCCGACGTGCGCCCTGACCCTTCAGCTGGG
5 CCGTCGCCAATGGCCCCATCGCCACAGCCTGATCTGCCGGAATCGAACGAGGCGATCAAGCT
GCTGCGCGCCGTCGTCCACTCCGCGGAGGTGCCGCCAGCTCAGGCGCCGGTCTCGGATGCGCC
GCGCTGTGTTTTTCTTTTCCCGGCCAGGGCGCCCAATACCCGAGCATGGCCCGCGACCTGGT
TCGAAACTGTCCCGACTTCGCCCTGCACCTGGACCCCTGCCTCGACCAGTTGGCCGAACCTGCT
TCCCGAAGATCCGCGTTGCATCCTGTTCCGCGATGGCCCCGCCGATCGGCTCGACCAGACGGC
20 CTACACTCAGCCGCTGCTCTTCTCCGTGTCCTACGCCTTGGCGCGCTGGTTGGGCGATTTCGG
CATTGCCCCGATGCGATGATCGGCCACAGCCTGGGCGAATACGTGGCGGCCTGCTTGCCCGG
GCTTTTCTCGCTGAGCGATGCCCTGCTGCTGGTGAGTGAACGCGGCCCGCTGATGGGCTCGGC
CGCGCGCGGAGCGATGCTGGCCGTCCCTTGCCCGAATGGGAACCTGGAGGAACGCCTGGAGCT
25 TCTGCCGACGACCGAATCAGCATCGCGGCGGTCAACACCGCCGAGAGCTGCGTCATCGCGGG
15 ACCCAGCGAGGCGATCGAGCGCTGCGCCCAGCGCTGGGCCGCGCAAGGCCTGACCTGTACGCC
GCTGCGCAGTCCCACGCCTTCCACTCCGCGATGATGGAGCCGATTGTGCAACCTTCGGCCA
TGCTTTGGCACGGGTCACTTTCGCGCCGCCGCGCGCGCTGGATCTCGAACCTCGACGGCAA
GCCGATCGATTCCGCGGCGGTGATGCAGCCCGACTATTGGGTGCGCCACCTGCGCCAACCGGT
CCGCTTTCACGAGGGAAGTCACTGCTTGGCCGAGGACACCCATGCTTGGGTGCAAGTGGG
20 TCCCGGCCGAACCCTGTCTCTTCTGTCGCGCCGCCACCCGGCCTACCGTCACCAGCCAATCGT
CAACCCCATGCGCCATGCAGTCGAGTCGACGGGCGACGTGCGCCGGTGGCGCCAAGCGCTGGG
CGAACTATGGCGGGCCGGCATGCCGGTGCCTGGGAGCGGCAGCGGCGCGGCCGGCATGCCGG
ACGACGTGTGCCGCTGCCGGGTACCCCTTCGAGCGGCGGCCCTTCGCGGCCCGAAGACCGGT
GGAGCTGGCGCAGCCCGCGCCCAAGGCGGAGCTGGTGAAAAACCCCGATCCCGCGCGGTGGCT
30 GTACCGCCGCGTCTGGCGCCCTGCCAGGCTGCGGCCGGCGGACTGGCGGTGCAGGCGACCGT
25 TCTGGTCTTCGGCGACGGGTCCGAGCTGTGCCGCGCGCGGTGCTCAGGTGCAGCGCCAGGG
GCTGAAGTGCGTCTCGATCACCGGGGCCGCAATTGCGCGGGGAGAGCGACATGCGCTTCAC
40 GCTTGACCCCGCTGATCCGCGCCAGCTCGACCAGCTCTTCGCGGCCCTCGATGGCTCAGGCTC
45 GCGGCCGCGGTACGTCTGCACCTGCTGACCCTGAACCCGCCCCCGGATGCCTCGGCGATCAT
30 CGCTCACAGCTACTACAGCCCGATGGCCTTGGCTCATGCCTTGGGCGCCACGAGATCGCGCC

5

87

10

15

20

25

30

35

40

45

50

55

TGTCTCGATCACCGTCGTACCGCCGGGGTCGTGCGCGTTCGCGGACGAAGCGATTTCGCGAGCC
GCTGCAGGCGCTGATCGTGGGCCCCGTGCCTGGTCATCCCCGAGGAGTTTCCCGGGCTCAGCGT
TCGGCTGCTGGACGTCAACGTCGACGATCCGGCACCGCGTCTGGCGGAGCGGCTCGTGGCCGA
GCTCTCGGGCACGGATCACATGGTGGCGCTGCGGGCGGCGAGCGCCTAGTGGCCGATGTCGA
5 TCAAGTCGATGGCCTCGGTGTGGGGATCGCCAAGGTGCCCTTGCGCCGCGAGGGCCACTACCT
GATTCTCGGCGGCCCTGGGCGATATCGGCTACCACTGTGCCCCGTATCTGGCCCAAACCTACCG
CGCCAAGCTGACGCTGACCGCGCGTTCGTCACTCCCGCCGCGCGCTCGTGGGAGCGAATGCT
GCGCGAGGGAAACCTGGATTCCCGGCAGCGCACGCGCATCGAGCGCGTGTGTGCTAGAGGC
GTGCGGGGCCGAAGTCCAGACGGCTGCGGTGCACTTGGGCGATCGCCATCGCTTGGCCGATGT
10 GTTCCGCGAAGCACGGGGCCGATTGCGGCCATCGCGGGCGTGATTCACTCGGCGGGGATTCC
GGGACACGTCCACTCGATCGACGAGCTGGTGCCTGCGCGACGAAGCCCAATTACCGCGAA
GGTTCGAGGGCTGCACCACTGGCCGAGGTGCTCGATCCGCTGAACCTCGACTTTTGTCTGCT
GTTCTCCTCGCTCTCGACCGTCTCGGCGGGCTCGGCTACGGCGCCTATGCAGCGGCCAACGC
CTACATGGACAGCTTCGCCCCGCCACGATCGGCCGGACGAATGTCGTTGGATCGCGGTCAA
15 CTGGGACGCCTGGCTGTTTGAAGCCAAGACGTGCTCGGTGCGCGCCGAATTGGCGCGCCTGGC
GATCGTGCCCCGAGGACGCTCCGGCCCTGTTGCGCGGGTGCTAGAGCGACTTCCGCAATCGTT
CATCGTGTCCACCGCCGACCTTCGGGCCCCGATCGACACTTGGATCCGGGACAAGAACCAGCGT
CCCGCCCCGCGAGATCCGAGCGGTTCAACCGCGACCGGACCTGAGCCAGGCGTACGCCCCGCC
GATCGGCCCCGTGGAGATTCAACTCTGCGGGCTGGTCTCCGCCTATTGCCGTTTCGACCGGAT
20 CGGGCGGGACGATTCTTCTTGAATAACGGCCTCAGCTCGTTGACTTGATCCAGCTCAGCTC
GCGCATTACCGCATCACCGGCAAGGATCTCAATACGACCCAACTGTTTCACTACCCACCGT
GCGCGCCTTGGCGCTCTTCTCGGCGGCAACCGAGGGGCTCGGCGCGGAGGAGCCCGCCAT
GGAGAACCTGTGGCTGCAACGAAGCGATGCGACCTCGATGACTGAGACCGAGGTGCGCGACT
GCGGCGCTACCGACCGCGGTGAGGATTTTCCGCGCAGCGATCCGGGACGACTCGCTGAAGAA
25 GCGCGATAGAAGAACGGAATCGTGTATGAAATACGAAACCACCGGATTGGAATTGGCCGTAT
CGGTCTCGCTTGCCGCTTTCCAGGCTCACCCGATCCCGAACAGTTCTGGTGAATCTGCGCGC
AGGTGCTCCGGAATCCGCCATTTTACGCGATGCCGAGCTGAGCCACATCCCGCATCCCTGCG
TCACCATCCGCATTACGTCAAGGCCAAAGGCGCGCTGGACCACGCCGATTTTGAACAGCCTT
CTTCGGCTACTCGCCCAAAGAGGCCGAGGTGATGGACCTCAATTCCGGCTGCTCCATGAGTG
30 CTGCTGGGAGGCGCTGGAGTCAGCGCGCTATGCGCCGAGCCAATTTCGCGGGTCGGATCGGCTT

5

88

10

5

15

20

10

25

15

30

35

20

40

25

45

30

50

55

GTTTCGCGGCGGCGGCTTCAACGACGGATGGATCGCCGGTACCCTCGACCGGCTGCGCACCGG
CGTGGGTTTGTAGCTCCCTGGAACCGCGTTCTTGACCCTGCGCGATTACCTGACCACCCAGAT
CTCCTATCGGCTCGATCTGCGGGGCCCCAGCCTGCTTGTCCAAACCGCCTGCTCGTCGTCGCT
GGTGGCGGTCCAGCTCGCCAGCAGGCGCTGATCTCCGGCGAATGCGCCCTGGCCTTGCGTGG
CGGCGTGTGCGCGACCGATCCGCTGCATTGCGGATACCTCTATGAACCCGGCAACATCTACGC
GCGCGACGGCGTCTGCCGACCGTTTCGACGAGGCAGGCGCCGGTACGGTCTTCGGCGACGGGTG
CGGCATGGTCTGCTCAAGCGGCTGAGCGACGCCCAGCGGACGCGGATACGATCTGGGCGGT
CATTGCGCGGGCGGGCGTGAACAACGACGGGCACCAAGGTTGGCTACACGGCTCTTGGCAC
GAGGGGCCAGGTGGCTTTGCTTAAAGTGTTTATCGCGCGAGCCGGGTGACCCGGCGACGCT
CGGCTACCTGGAGGCCCCATGGCACCGGCACCGCGCTCGGCGATCCAATCGAGGTCGAGGCGCT
TACCCAGGCCTTCGCCAGCAAACGTCGCGGCACCTGCGGCTTGGGCTCGGTCAAGGGCAACCT
GGGTCACTCAACACGGCGGCCGCGCATCGCTGGACTGATCAAGGTGGTGTGCTGGCGCTGAAACA
TCGCGAAGTGCCACCCACCTCAATCTGCGCCGTCCCAATCCGAAAATCCGCTTCGACGAGAC
GCCGTTTTTCCAGTCGTCGAGTTGCAACCTGGCCAAGCGGGACCGGCCCTTGCGAGCCGG
CGTGAGCTCCTTCGGCATCGGCGGTACGAACGCCCACGTATCTCGAGGAGGCACCGCCGAC
GGCCAACCCGGCGCCACACGGCAGATTCCGACTGTTGCCGCTTTCGGCCAAGACACCGGCTGC
GCTCGAAGCGAAGCGCCGCGATCTGGCCGGCTTCCTCGAACGCCACCCGGAGACCTCCTTGGC
CGACCTCGCCTTTACCTGCAACGCGGCCGCGAGGTCTTCAGTCACCGCGCCTGCCTCGCCGT
GGAGACCTTAACGTCCGCGCGCACGCGGCTGAGCGGCGAGTCGTCGAGCACTTGGCTGGTGGG
CCCCGCGCCCAGCGCCATATTTCTGTTCCCTGGTCAAGGCAGCCAGCTCGCCGGGATGGGCCG
CGGTCTGTATCACCATTTTCGAGCCGTTCCGCACGGCCGTCGATGCCTGTCTGCGCGAGCTGGA
GCCAGGACTGCGGCAAGCGCTCAGCGCCCATTTTCGATCCGAATCGCGGCGCGGACCCACCCGA
TTCGACGACCTTCGTCCAACCCTTGTGTTCCCTCGTCGAGTACGGGGTGACCGAGTGGCTACG
CTGCTTGGGTGTGCGGCCAACAAATGGTGTGGGTACAGCTCTGGCGAGTATGCCGACGCTG
CGTCGCGGGCGTTCTGTGCGCGTCCGCGGCGGTCTCGCTGCTGGCCGAGCGCGAGCGGCTGCT
GCGCGACCTGCCAGCCGGCGCCATGCTCGGCGTCCCGCTGGCCGCGGAGGCGCTCGAGGCGAT
GTTGCCCGACGCTCTCGATCTGGCGGCGATCAACGGCTGTCAGCTTTGCGCCGTGTCCGGGCC
GGTCGCGGCGGTCCACGCCTTCAAGGCCCAACTGGAAGCCGCCGACATCACGCCCGCCTGTT
GCACACCGATCGCGCCTTCCACTCGCGGCTGGTAGCACCGGTGCTTGACCGGTTCCAGGCAGC
CGTTCAACACGTGGAGCTGCGGCGGCCGCAAGTACCTTACCTCTCGACCGTCAGCGGGCGATT

5

89

10

5

15

20

10

25

15

30

35

40

25

45

30

50

55

GGAGGCGGATGGGCCGGCGAACCCGCACTACTGGGTGCGTCACCTGCGCGACACGGTGCGGTT
TGGTCCAGCCCTGGAGGCGCTGCCGCCGGTGGATTCTTTCGTGTGCATCGAGGTGGGACCAGG
CTCGGCCTTGAGCACCATTGGCGCGCGAAACGTTGGGTTCACAGGCGCGACTGATTTCTGTGCT
GCCGCGGCCGCGAACGGGGCAAATCGAGCCCGGTCCGGTATTCGAACGACTGGCGGCGCTTTG
GCGCAGCGGGTTGACATTGGATTGGTCTAAATTGACGGGCGGCGAAGAGGGTCATCGAATTCC
CTTGCCAGTCTACCCGTTTCAGCGCAGCCATCTGTCGAGCTCCCTGGCGGCGGGCCACACGCC
TTCGTGCGGGCCTGCAGTCGAATCAGGCGCCATCCTTGCCGAGCGATCCGAGGGGAAAACGC
TGAAACCCGGGATTGCCCGCTGCCAACCGCCACGCTCGAGCCCAAGCGGTGCTCCGGCCCC
ACTCGAGGCTACCGACGCGCGCAGGTAATCGCGAGCGACTGGCCGAACCTTTGGCGCGAGTTGCT
AGGGTTGACCTCGATTGGGCCCGACGACCATTTCTTCGACCTGGGCGGCCACTCGCTGACCGC
CACGCGGCTGCGCGCCCTGATTACACAGCGGTTTCGATGTGATCTCGGGCTCGACGAAATCTT
CGCTCATTCGCGTCTCTCCAGCTGGCGCGCCGTATCGAGGCGGCGCCAAGAGCCGATTTTC
CTCCATTCCCAGCGCGCCGGACCAGGACGACTATCCCTTGTCATCCGCCAGCAGCGGATTCA
CAGCATCGTCACGAGGGCCGAGGTCCGCACTGCTTATAATTTCCGATCGTCTCTGAGCTGCA
GGGCGCTCTGGATCGAGTGCATTTCGAGGCGACGTTTCGCGGCATTGTTCCGGCGTCATGAGGG
GTTCCGCACCCGCTTTGTGATGCGCGATGGCGGGCCGCGCCAGCGCATTGTACCGGACGTGGC
GTTTCGCTGCGCTCACCCAGGTCGAGCCAGAGCAGGTTCCCGGGCGCATCGAGGCCTTCAT
CCGTCCCTTCGATTTGGAACGCGCGCCGCTGTTCCGCGCGGAGCTGTTGCAGTTGGCCGAGCA
GCGCCATCTGCTACTTTTCGACATGCACAACTTAATTGCCGACGGTATCTCGCTCAACCTGTT
CGTCGCCGATTTTCGCGGCCCTGTACCATGGTCGTCCGCTGGCGCCGCTGAAACTCCGCTATCG
CGACTATGCCGTTTGGCAAGAGGCGCGGCTGGCCTCCGATGACCTGCGCAGCCAGCGCGAATG
GTGGCACCGGCGGCTTTCCGCCCGGTCGCCACGCTGGCGCTCCCTCCCGATTTCGCGGTCC
GGCGGTGCGCCGCTACAAGGGCCGTAATGTGGTGTTCACCTGGACCGGGAGATCCGCGACCG
CCTGGTGGCCCTGGCTCGAACCAGGGGGTCACCATGAACGTGATGATGCTGGCGCTCTGGGC
TGCGCTGCTGCATCGCGAAACCGCCCAATCGGAGCTGGTGGTCGGATCGCTGCTCGGCGGGCG
GCCGCACAGCGAGCTGCATCCCGTGATCGGGCTCTTCACCAACTTTTGGCCCTTGCGGTTGGC
GGTCGAGGGATCGACCCGCTTCGATCGCTTCCTTGCCGCTTGCCACCAGGTGTTTCTCGAAGC
CTATCAGCGCCAGGACTATCCGTTCCACTTGTTAGTCCAGGAACTCGTGCCGGTCAGGGACCC
GTCGCGGTGCGCGCTGTTCCAGACCTCGCTCGTCTACCACAACGAAATTGACGGCAAGACCAA
GCTGGAATTGGAAGGGCTGAAAGTCGAAGTGTTCCCTTCGAAAAGGTGTGGCGAGGCTGGA

5

90

10

15

20

25

30

35

40

45

50

55

TTTGAAGCTGGATGTGACACCTTTTTCCGACCGACTCGAATGTGTTTTGCAATACGACTTGGA
TCTGTTCTGCGAGGAGACGATGCGCGGCCTGATCGCGCGGTTCAGGCGTTGGTGGCGGGGCT
TGTCGCGCGATCCGGCGCAATCGCTCGCCGCCGCGAGCGTTTCCGGGAAGCGGGCGCTGCGCGC
GGGCGTGGCCACGGCAAGCGAATCGTCGCCGAGTCACTGCCGCCGAACCATCGACGGCGTA
5 CGCCACTCCCTCACCAGTCAACCGTCGCGGCTAGTCTGACGGGACCCGCCGACCTGCCCGC
15 GATCTTGGCGGCCTACGTGGGGCAGAACCCCATCCGTTGCGGATCCATCGGGGTCTCATTTT
GGAGGCGCCGCTGGGGTTGCGAGCGCTGCGGTGCGCGCTGGACGAGTGTCTGGAGAACAAC
CCATTGGCGCAGCGTGCCTGCGGGCGATCGCGCGCGGCGGTGGATAAGTTGGAATTGACCAG
CCTGGTGGCGCTCGACGACCTGCGCGGGTTGGTCAATCCTCAGGCGAATGCCTTCACCCTGGC
20 TTGGCGCGATCTGGCGATGCCGTTCCGGGAGGGGCGTCCCCTGTGGCGACTCCGCTGGCGTG
10 GTCGGTCCATCGCGCTGGTTGCTATTGCTGACGTTTCAATTCATTGATCGGCGACAACGGCAC
GGTCGACCTCTTTCTGGCGGCACTCGCCGATCACCTGCGCGCGCGTCCGCTTTTCCCGTAGC
25 ACCGCTCGATGAGGCCGAGCTGGAGGCGGAGCTGAAGTGGGGAGAGGAAGGGGAGGGCCTCGG
GCTGACCGCGATCGCGCCGGTCTGGGCCAATTGCGCGAAAGTCCGCTGAGTCTGTGGCCCA
15 GATGTGGCTGGACGAGGTCTGTGCGCGCCACGACCTCACCCCGCTAGAGGTCTTGGCGGCCCCG
GCTCCTCGATTGGACACGAAGCCACGGTCACGGGTGATCGCTTTGTGGACGCCGCTGCCCGA
30 GGACCATCCGCTTCGCGATGAAGGCCGCTGCCCTCCAGGTTGCGCTGCTGGAGGGGCCCGCTC
GCAGCGAGGAGCGGGCGATCCAAGCTGGCTCGAGCAAATCGCCTTGAGACGGGGTACCCCTGC
AACGGAGGTGCTTTGCCCTACTCCGACCCAACGGGCAGCCATCGACCTCGCGCTGGCCTGGCT
20 GCGCGAGCCGCTCTTACGGTTTGGTTCGGAACCGTTCAGCCGTGGCCGGAATCTCCATTGGT
35 CTGTCCGTTTCCCTCAATCTCGCGTTCCGGCCAAGCCATCCAATTGCCTACGCGCTCAAGCA
CGAGGCCACGCTCGCGGTACGGCACGGGCGCGGATCTGATGCGTTTCTCGACGGCTTGGG
CCCCGAAAGCTGAAGATTAGCATAAGCGCCCGGCCAAGGGCATCCTAGGATGACGCAAGCCTC
40 GGCCGCGTCGACGTCCCAGGTGCGCGCGGAGGTCACCCCGGCCGAAAGACGACGATGACGA
25 TCAAATCCGAGATGTCGGCCGTTGCTCACTCTGCGGAGAGCGGCTTCCGCGCTGGGCCACGCG
TGGGCGGCGCGATGAAGCGGGGCCGACGCCGAGCAGGCCGCGTGAAGCTGCTCCGCGCCC
CGGTGAAGCGGAAGTGGCTGCCCCCGGCGCCCTGCGCCTGAGCGAGCGGCGTATCCCGG
45 AGGTGTGGGCAGGCTACCGCGGAGCGCGGGATGACCCGAGCCCCCGCCGCGCGGACCAT
GACGCCGCCCCACGGGGCGAGTCGTCCGGCGCGCGGCGCGCTCGGGGCTTCCGCCCGGGG
30 CGGGCAGGTGCAGGATGGTCCGGCATGGTGACGCTCCGACGTCCGACGGCATCGAGGACGAG

5

91

10

5

15

20

10

25

15

30

35

40

25

45

CTCGCGCCGTTCCCCCGGTCTGCGCGGCTGGCTCATCGAGGGCGAGCTCGGCCGCGCGGG
ATGGGGCGGGTGTTCGGGCGCGGCACCCGAAGACGCGGGCGGGCGGCGATCAAGGTGCTG
CTCGGCGACTACGCCCCGCGGCGGACGTGGTGGCCCCGCTTCGGGCAGGAGGCGATCGCCGTC
AACATCATCAACCACCCGGGAATCGTCCGCGTCTTCGACTCCGGCGAGCTCGAGGACGGCTCG
CCCTACATCGTGATGGAGTACCTGGACGGCCGGGGGCTGCGCGACTGGGTGCAGGCCGTGCCG
CCCCGCGAGCGGCCGCGGCAGGTCTGTGCGGCTCGGCTACCAGATCGCCTCGGCCATGGCCGCG
GCGCACGCGTCCAAGGTCTGTCACCCGATCTGAAGCCGGAGAACATCATGGTGGTCCAGGAC
GAGCTCGCGCCCCGGGGCAGCCGCGTCAAGATCCTCGATTTCCGGCATCGCGAAGGTCTCTGG
GGAGGTCTGCCCGAGGTGCTGGAGCTCGAGGGGAGAGGCTCCCTCGCGCCCCGCTCCGCGTCC
ACGATCCGCACCGAGCTCTCGACGCGGCCGCGCCGACGGTGGGCGCCACGACCGGCCAGAG
AGCCCCGCTGGGCGCGAGCGCCACGCCAGAGAGCGCCCTGGGCGCGAGCGCCACGCCAGAGAGC
GCCCTGGGCGCGAGCGCCACGCCAGAGAGCGAGGCCACGAGGAAGACGCGCTCCGGAGCCTC
CCCGTCTGTACCAGCGGCAGGCCCGCGATCCACCCGCGCCGGTCTGAGATCCCGCCCCGAGGCG
GTCTCCTCCGCGGCGTCTGCGCGGGTCTGCGCGCGTCTGATCGAGCCAGGCGCGCCCGCGCCGAG
AGCGAGGGCGCGGACAGCCCACGATGCCGTTACGCAAGAGGGCGTGTGGGGCTCGGGACG
AGGAGCTACATGGCGCCGGAGCAGGAGCGCCACTCCGGGAGCCTGGACGTGAAGGCGGATGTC
TACTCGCTCGGCGTCATCTCTATGAGCTGCTCGAGGGGCGACGCCCCGACGCGCCGAGCGCC
GCGTGGCCGCCCCGATGAGCGCCGCCACGCCGCCGATCTCGTCCGCCCTCGTCCACCGGTT
CTGGCGTTCGATCCCAGTGC GCGGCCGCGCATGGCGGAGGTGGCGAGCGCGCTTACCGGCTC
GGCCGGGCGAAGAAGGAGCTCGACGAGGCGCTCTCGAGGTGGGTCTGTCGGCGAGGGGCGCCG
GGGCTCTTGCCGTGCGGCTATGCTCTTCTCGAACTGGTCCTCCTGGGCCCTGGGAATTATAC
GATTCTTTCCAGCCTGTAAGTGCATTTTTCTTTCAATATCGTCTCTCTTCATATACGAGGTG
AGTTCTCTGAGGTCTCTATAAGTCTGGGGTGTCTATTCCGGCTCTTACTTGTTACTTCGC
CTTCTTAGGAGTTTTTCTTAATTTTGCCCTCTTACATTCCCGTATTCAATTCTAACTGGGCCC
TATCTCATTCGCTAATACGTTTCTGTATTGTGTACATCTCCTATCATGTGTCAATACTTGTTT
CTGTTTATCATTATTCTTATTGTTTACGCTCTTATTTCAATCATAGTATAACATTAGTTTACT
GATTATCGCACTTGAATTCGCG

or its complementary strand,

30

50

55

5

92

10

(b) DNA-sequences which hybridise under stringent conditions to regions of DNA-sequences according to (a) encoding proteins or to fragments of said DNA-sequences,

15

5 (c) DNA-sequences which hybridise to the DNA-sequences according to (a) and (b) because of a degeneration of the genetic code,

20

10 (d) allele variations and mutants resulting by substitution, insertion or deletion of nucleotides or inversion of nucleotide segments of DNA-sequences according to (a) to (c), wherein the variations and mutants offer isofunctional expression products.

25

8. DNA sequence according to claim 6 selected from the following

30

(a) open reading frames:

<u>Nucleotide Position</u>						
35	20	ORF1	1666	-	1	Seq ID No 3
		ORF2	1605	-	3338	Seq ID No 4
		ORF3	6100	-	3398	Seq ID No 5
		ORF4	7110	-	6374	Seq ID No 6
40		ORF5	9590	-	8433	Seq ID No 7
	25	ORF6	11393	-	9855	Seq ID No 8
		ORF7	13656	-	12712	Seq ID No 9
		ORF8	15374	-	18984	Seq ID No 10
45		ORF9	20003	-	27889	Seq ID No 11
		ORF10	28251	-	29402	Seq ID No 12
	30	ORF11	31720	-	30401	Seq ID No 13

50

55

5

93

10

ORF12	31982	-	32932	Seq ID No 14
ORF13	33128	-	33613	Seq ID No 15
ORF14	33661	-	34007	Seq ID No 16
ORF15	35611	-	35255	Seq ID No 17
5 ORF16	37856	-	35730	Seq ID No 18

15

or DNA sequences complementary to said open reading frames,

20

(b) DNA-sequences which hybridise under stringent conditions to regions of DNA sequences according to (a) encoding proteins or to fragments of said DNA sequences,

25

(c) DNA-sequences which hybridise to the DNA-sequences according to (a) and (b) because of a degeneration of the genetic code,

30

(d) allele variations and mutants resulting by substitution, insertion or deletion of nucleotides or inversion of nucleotide segments of DNA-sequences according to (a) to (c), wherein the variations and mutants offer isofunctional expression products,

35

and peptide sequences corresponding to said open reading frames

40

SEQ ID No 19 (>ORF1)

25 VDPEREAVTLGLAFNRAQGRITYARGPEARA EYIGTAMRAADVIEDRFEIERLAVSGGMGDVYR
ARDRVSGQAVALKVLQASANDLRRFAEAEALVTLRLPGVVQYVAHGVTGAGRPYLAMEWLD
GVTL EERLAGAPLTLAESVALAARVATT LGAIHWLGVVHRDLKPSNMLVGGAVERTLLDFG
45 IARHLRLAPTLTSPGAVLGTGPIAEPQVRGDAPVDARDVFALGCVLFQCLAGRPPFLGNSAL
ALLMRVVLEEP PRLGELRDGIPEPLERLVARMLAKNAGERPRDGAAAAAE LAAVAGEGLSIGA
30 SAVAAPAAPGEAITTAERKVMCVILAEDGGAEAGATLSEDDGAARAEALRDIAARHGGRDLRL

50

55

5

94

10

QARWWLVALSGAESPTDLATRAAHCALALRAALGGVPVSVATGLAEVEARLPVGELVDRVAQL
IAGRDI.SPPEIRLDDATASLLASRFETVQGGGCWLRGPKEEPDAVPRLLGKPTPCVGRERE
LSQLATEWRHCVDEPSANAVVVVGAPGLGKSRLAWEFLLRTLEQREGAAI

15

5 SEQ ID No 20 (>ORF2)

20

VRPCARLNASPSVTASRSGSTAAGSVHASTSACVEQPATGRTQPASPRWPPGAAALRLTSAMP
RWFNTAGPCNPADHYMLPAEERLPAVRDLVDRKAYFVLHAPRQIGKTTSLRTLAQDLTAEGRY
VAVLVSAEVGAPFSDDPGAAELAMLAEWRTAGAQLPADLRPPPPDAPAGQRIGAAALRAWAQ
AAPRPLVVFLDEADALRDATLVSLLRQIRSGYPDRPRDFPHALALVGLRDVRDYKVASVDSGR
10 LGTSSPFNIKVESLTLRNFTREVAITLYAQHTAETGQVFRPDAVDRAFELTQGGPWLANALAR
QLVEVLVKDRAQPITSANVDRAKEILIERQDTHLDSLVDRLREPRIRAVIEPMLAGTALPSVP
PDDLRFALDGLVRMTAEGGLDVANPIYREIIVRELAFPIRASLPQIKATWLTQDGRLLADRL
25 LDAFLSFWRQHGEPLLGAAPYHEIAPHLVVMAFLHRVVNGGGTVEREYAIGRGRMDLCVRYAG
ETLAIELKVWRDGRPDPAEGLAQLDYLAGLGLDRGWLILFDQRSQGPPIAERTRRERALS
15 AGREVAVIRA

30

SEQ ID No 21 (>ORF3)

35

VTIKKTFRSIDPATLPKHFDSPVAELRLADLWEADGTYRYDPSRPREETFVVDTPPPTASGSL
HIGHVFSYTHTDVVVRQRRMRGNIFYPMGWDDNGLPTERRVQNYFHVRTDVRTPYERGLTLP
20 QAAPETIKKEPPRIVSRPNFIELCHKV TREDEQVFKALFRRVGLSVDWRNEYATIDDHCRRTA
QLSFLDLHEKGHLYSVFAPTMWDVDFQTAVAQAEVEDRPQSGAFHDI AFAVEGTAEELVIATT
RPELLAACVGVT AHPEDPRYQHLFGKTALTPIFRAPVPIFPSPLVDREKGTGILMVCTFGDAT
DVIWWREQKLPLRQMLGKNRVLPTVTFGEGAWESRDPAAANAAYAPLQGRGVKQARA VVELL
40 RREEHAAAPGRGPALRGEPRPIERAVKFYERGDQPLEFVPTRQWFVRLADKKAELLEYGDKIK
25 WHPDFMRLRYRNWTEGLQGDWCISRQRYFGVQFPVWYPLDAEGNPDHSRPLLATREMLPVDPT
VDVPPGYEASQRDQPGGFTAESDVFDTWFTSSLTPISSHWDGDDPARHARLFPADLRPQANDI
IRTWAFYTIKAMLHESSVPWHHVAISGWILDPDRKKMSKSGNVVTPMHLLDTYSSDAVRYW
45 SASARLGTDTAFDEKVLKIGKRLVTKIWNASKYVLSQSAEVHPISEELDRALLHKLSAVVDDA
TRSFDEHEFAAALERTEDFFWRWFTDAYLELAKARARGE GAGEAARGSAVAALRLGLSVLLR
30 LFAPVLPYITDEVWRWVYAEETGDTSIHRAKWPSAADFAAVAAPSDPGLLDLAAAAA VVKNR

50

55

5

95

10

KSELGASVGRVVTDLALGANAATLARKPALGDVLTAVRAGAHALVRPELADGEVLVVRCELE
PAAAAAAGAGGAAASEE

15

SEQ ID No 22 (>ORF4)

5 MIHAEPFEARLVAAARPLSPFVRELSFERADGRSFLFEAGQWVNLVLPPLPGGEVKRAYSIASAP
DGSPRFDLAVTLVQGGAGSEHLHRLPEGATLRAIGPHGLFTRDPGDSAPSLFVATGTGITPLR
SMLRASLRAGLAAPHLWILFGARFEEDVIYRDELEALARGSDRIRYEITLSRGGPSWAGRRGY
VQAHVPELYRELAEKSGDPAPHVFICGLDRMVSSVRELARGELGVHRKHVHVERYD

20

10 SEQ ID No 23 (>ORF5)

MKSLPSDRAARLAQSDIRMTLACAXVHGINMSQGVCDTPVPSVILQAVKEAMDRGCNTYSRF
DGIVELRHAIAAKLARHNGIAADPETDITVSAGATGAFQATCMALLNPGDEVLLFEPFYAYHA
25 QAILAVEAVPRYVTARSLSWNVGDDELERAITPKTKAIVVNSPGNPSGKVFGRMELEQIADLA
CHHDLMVITDEIYEYFIFDGREHVSVASLPRMSERTITIGGYSKTFSTITGWRIGYSVADARWA
15 KAIGAMSDLLYVCAPTPLQHGVAAGIRGLPRSFTYTGLAQGYERKRDRFCRALEKAGLPFCVPQ
GTYTVLADVSRLPGRTGRERAIYLLDETGVAGVPGDAFFEGTQGSRFMRFCFAKTDEDLEEAC
30 QRIEQLA

30

SEQ ID No 24 (>ORF6)

35

20 VSDPRKERLGMDLEEFRRIGMRIIDWAADYLGHDPDRYPVFPAIRPGDVKGRLAPTPPVEPEP
MDAVLTDFEQIILPGITHWNHPRFFAYFANTASGPGILGELLAACLNVNVMWLWRTSPAATELE
ELVLSWLRQMLDL DAGLHGAIMDTASTASMVAIAAARDSAEPTIRLRGMAGQRRMRLYASEQA
HSSIEKAAITLGIGQEGVRKIPTDFAFRMVPEALRAAVVEDLGAGLRPFCAATVGTSTTSV
40 DPIP AIVSVCREHGLWLHVDAAYAGMAAIVPEHRDVLAGEGADSLVNVPHKWLFTPMDCSVL
25 YVRDADRLKRAFSLVPEYLRTEGDVTNYMDWGIQLGRRFRALKLWMIVRYFGHEGLAARIREH
LRIGQQLAQWVDADPDWERLAPTPFSTVCFMRPSALACIMRSADEAERESIERELDRLNEAL
LDEVNKSGRVFLSHTRLHGRYTIRVAIGNIRSDEVAVREAWECLEAAGARLCADERFVSCSR
45 ADEGRGKS

45

50

55

5

96

SEQ ID No 25 (>ORF7)

10

5

15

MRREEPVLEAFYERYCAAPRETSYHVELPVDVELHQEAAPALPOARSLELAGRVALVTGSSRG
IGKAIALRLAEQGADVAVNYHSNKDAAEQTAAEIRALGRRTMVVQADVTRPNAAAELFSSVEA
QLGPIDILVNNVGDDFFKPLAAMTDDDEWRNVMDSNLSSVHYLCRAAVARMRQRKSGRIINIGL
SPTYAIRGAPNVAAYSIAKTGVLILTRSLATEEAPHGILVNCVSPGLIDNGYLPQAQKEWMER
RVPMGRLGRASEVADAVAFLASDRASYVSGANIAVAGGWDWTDGRGTEHRRVDLFIGHEEP

SEQ ID No 26 (>ORF8)

20

10

25

15

30

35

40

25

45

30

50

55

MSGRFPGARNVEELWQKL RAGVECVVTFTEAEALAAGVSREMLANPSYVRRGAPLDGVELFDA
SFFGFSPREAESMDPQQRIFLEVAWEALERAGYDPAHSGPIGVFAGSAPSGYHSLAQSDPEI
LGALGHYQLTLNNDKDYLTTHASYKLNLRGPSVCVQTSCTSLVAVVMACQSLNHECDMALA
GGVGIIHAHQRRGYLYQENGISSPDGHCRAFDVAAKGTVGSGIGIVVLKRLADALADGDHVHA
VIRGAAINNDGSSKIGYTAPSVQQAQAEVIGMAQALAGVEPDDISYIEAHGTGTPLGDPPIEIAA
LTRVFRAKTARRQFCAIGSLKTNLGHLDAAAGVASLIKTVMALHRELPPSLHFERPNPKLEL
ESSPFYVNTRLTPWHAARGPRRAGVSSFGIGGTNAHVLEEAPAPPPSGPSRRWQLTLAARS
EAGLARATADMI EHLDRHSGTSIADVITYTSHVGRRAWPFRRVVGESAADLRAALASEGSPRS
ISSCQAARERPVVFLFPQGAQHLMARELYEVEPIFRQSLDRCAELLRGPLGLDLRQVLYPA
EGQRDDAEQELGRTAIAQPALFAIELSLAKLWMAWGIVPQAMIGHSVGEFAAACLAGIFREED
ALRLVAERGRLMQMPPGAMLAVPLAEPAPYLSDDISLAAINGPALS VVAGPIE AIDALAA
ELLDHGLSCRR LHTRHAFH SKMMA PVVDAFTRCVSAVERRRPPSGHFLSTLTGGWISPEAATIP
AYWARQLVEPVRFQAQAVRQLLSESTWLWLELPGQTLSPVLRQQARADGGQVVVASLPRAKDA
GADHLAVIEALGRVWSAGGTVDWKRFEHGEARRRVLLPTYPFERQRYWASPRHTSAPPEAIIK
PLLAKNPVADWFFLPARRSDPPVSFDAQAVTTRSTWLVFIGDEGLGAALVEGLARRGHEV
VAVVTGERFEQTGTQRYTIDPAANGDVASLFARLEIEGRMPDRIVHAFCTSPADGARIERGAA
LEIERRLGFD SLLLLAQVIAAQRHPKPLMLGVITTRAHSVIGTEIIIEPLRALVLGPCRVIPQE
IPHVSCRNIDIDLPGEGGRAEIAARLIADLERESPDSVVAYRGRRWVESIELTDVGRRSAGA
APRLRQRGAYLITGGLGGIGLVAAELLAREAHARLILVGRGTGLPARQGDDWLAAHGAGDATS
RKILRIRALEEAGAEVKIAAADVSDFNAMRSVIEEARTFRIGRIDGVIHSAGIASGGMIQLRTP
MAAHRVMA PKVGGTLVLDALLRDERPD FLLICSSLASLVGGATQIDYCAANAFDAYAQSREG
EEGCRVISVQWDTWSDVGMVDFKLPADLQEGRRRESLKRGISSEGAEVLGRILSAGMSGPLA

5

97

10

ICTSDLPAYKQSVTTRRSQHEQTPAARPMHSRPTTTGAYVAPETETERRIAAIWQDLLGLEQV
CANDDFLQLGGHSLLATQVLSRVLQTLKVGISLPQFFDAPTVAGLSRLVDAARAEGAGPVAPA
IGRVERDAYRIKPPAAEQAARTKP

15

5 SEQ ID No 27 (>ORF9)

20

10

25

30

35

40

25

45

50

55

MEPVGGVDMNQPAKQQETCVFPTSFAQRRLWFLDQLEPGSAVYNMPASFRTGRGPYDVDSLVR
VNEIVRRHESLRRTTVDIVIDGEPVQVIAPSLRIEVPVVDLSEIDEPERAEARRLMAEESRRPF
DLTRGPLLRKALLRLGEADHVLILTMHHIVSDGWSMDVLFKELSTLYAAFHEGRPSPLPELPI
QYADFAVWQRELLQGEVLESHLGYWREHLRGAPTLLLELPMDRPRPPAQTFRGSQRAFRPLPSL
QQAVQALSREQGATPFMTLLTAFSVLLSRYARQSDLVVGTPIANRTAELEGLIGFFVNMLAL
RIDLGGDPSFRELLGRVREVTLGAYAHQDLPPERLVEELSPGRSPSHSPLFQVSFTLQNTPM
ATNRADIASGGAPLVEMKAAKFDLILELSESPQGLLGTFEYNTDLFDAGTIERMAGHLEVLLS
SAVAAPDRPIAELPLMGAERSRVLVEWNSTAALYPEDHCMHELFEQQVERSPEATAVLLQQQ
TLTYRELNMRANQLAHLRLSLGVGPEVRVGLYLERSIETVVAAILGVLKAGGAYVPLDPTYPSE
RLGLMMADAAPSVLLTQASLLSKLPPHGDLTLVQLDALHEALSRLPHHTPRSGVTAQNLAYVM
YTSGSTGRPKGVLEHRLCNLPTVQAKLYGIAPGDRLLQFAPLCFDTSFCEIALALLSGATL
VMGTADELLPGPPLVELLKKHAVTAMLLAPTVLALPEQQSAALPLRVLTMAEACPAELVKR
WKAPGRRLFNISYGPTETTIWASSAADLSDERIPPIGRPIANTQIYVLDEALEPVPIGVPGEIF
IGGVGVARGYHGRPDLTAEFVDPDPFGQTKGARLYRTGDRARWLPDGNLEFLGRNDEQVKVRG
VRIELEEIIRAALLKHPAVAQAVVREDTPGDKRLVAYVVGRGGARVTAAELRQSVSERLPAT
MVPSSFVALDALPLTPNGKVDRLALPEPEQSAGGEDHVAPRNAVEEELARIWASVLRRLERV
HDNFFEIGGDSILSIQIVVRAQQAGLRLTPROMFQHTIAELSTVARAVEAVHVEQDPVTGPA
PLTPVQRWWLEQEAAPHFNQSI FLEVRERLDESALEQAI AHLIDHHDALRLRLARDERGAH
QVFAAPGGSTPFQVRDLGALPSAEQISAMEKAASEAQASDLAAGPVVRVAVLFDLGEVAPQRL
LVIAHHIAVDSVSWRILLDDLFGAYEQARRGEAVRLPPKTTSVKRWAELLTEHAGSEAVKAEL
GYWLDSSRRTVAPLPVDRRAGEDVWGSARHIVVSLTPEQTEQLLREVPQAYRTRIDALLTAF
AQAIARWTGSPAVLLDLEGHGREELAGVDLTRTVGWFTAMYPILLRVDAADPGEALKSIKEQL
RAVPGRGLGYGLRLYLRSDTIAEVRALPQAEFCFNYLGQLDQAIPEAAPFRPAREYQGSERSP
GAHRAHLIEVNASIANGRLYATWTYSERRHEPETIERVAASFVTALRALIAHCTLPEVGGNTP
SDFDKVRLRQETIDALDAIDAGPGPSARGSRIEDVYPLSPLQEGILFHTLYATDYTAYVEQFH

5

98

10

15

20

25

30

WTLEGDFDAEAFTRALQDVVARHAALRTSFAWERLDAPLQIVRTGAVLPVEHQDLRGLAAEEQ
TAHISRYVEAERQRRFDLRKAPLMRAGLLRLRKDAWCLVETIHHLILDGWSTQILLKEVFPLY
EAHRGHRCHLAELEQPRPYGDYIGWLAKQDQVRTAAFWRRELEGFSAPTPLGVDRAPPHDDG
GPRFGWRRIALSGDDAARLAAFARQHQLTMSTLVQGAWALLLSRYSGDPDVLFGMTVSGRSAP
5 IPGIERMTGLFINTIPVRVREPADASVLAWLKALQEHEAEELLEHEHSPLVEVQAHS DVPRGTP
LFESLVVFENYPVQVIFEAPPVEGPTRAEEGLRMIDAQYISDPPYPLTVVAAFHGTLYLNIGY
ERRRFDDQAVERMIGHVTLLRGFVQRPETSVRDLPLLTAAEEERTQLHAWNATAAPYPEGHCHM
HELFEQQVERSPEATAVLLQQQTLTYRELNIRANQLAHHLRSLGVGPEVRVGLCLERSIETVV
AILGVLKAGGVYVPLDPTYPSERLGLMMEDAAPSVLLTQTSLLSKLPPHGDATLVQLDALHEA
20 10 LSRLPHHTPRSGVTAQNLAYVMYTSGSTGRPKGVLVEHRGLCNLPTVQAKLYAIAPS DRLLQF
APLCFDTSFCEIALALLSGATLVMGTADELLPGPPLVELLKKHAVTAMLLAPS VLAALPEQQS
AALPLRVLAMAGEACPAELVKRWKAPGRRLFNSYGPTETTTIWASSAADLSDERIPPIGRPIAN
TQIYVLDEALEPVPVIGVPGEIFIGGVGVARGYHGRPDLTAEFVDPDPFGQTKGARLYRTGDRA
RWLPDGNLEFLGRNDEQVKVRGIRIELEEIRAALLKHPAVAQAVAVVREDAPGDKRLVAYVVG
15 RGGARLTAAELRQSVSERLPATMVPSSFVALDALPLTPNGKVDRRALPEPERSAGGEDHVAPR
NAIEEELTRIADVLGAKRVGVHDNFFDLGGHSLLLVRVHDLRGQRFRPPSMVDLFTYPTVA
SLARFLGERANGKQSPREAAADVTERGRRRLEARARRAKAIRGPT

35

40

SEQ ID No 28 (>ORF10)
20 MKHNIGWLLPAALATLAFVPACSPNHGEDAPSVTSAESGAAPSADCVALGAKLQAALDGAAAA
QKAPGAAAQVSGDCVWRGATGVSDLVASTPTKPGDLFRIGSITKTFVSTLILMLRAEGRSL
DDAVSKYVKGIPAGDQMTLRQILGHTSGLFDYTYSPALGQMIEVDPTRAFAPAELIALATAEA
PYFAPGAGFRYSNTNYIVAGLVAEAVSGGTLAGLLRTRIILDPVGLAHTYLDGAEPVQGLIRG
YG DYAGLVDITDQLSPTEAWAAGALVSNVDDLNRFFALLISHELLSSDELQDMTTWTPTMWP
40 25 HEPGYGLGLIERDSALGSLNGHCGIIWGFQASASYGVPGRGDAITALINRSDGDAARLVDELAK
VVKER.

45

SEQ ID No 29 (>ORF11)
MSIDRAVLEQLDRVGGRLAEGKALKLLEDIAWPREVEERFFAAGEDRLPEVEYRVDRDGLARR
30 VAELRELLCAIDGDAPALGWLRDNVRAQIQAAELLEAAGTRAFSARSQELYGCARSRFFGGSL

50

55

5

99

10

RNIDLAEHLTERLRVHGWDEASDPEEEPLDAGALRDMLAARVAGRAPRLDLEITVDPRVTAKV
VAGMSRVIRPEATFAAWEAEGLWHHEVETHALTAHNGAAQPRCAFLRSGGPRTTRTQEGLAI
FAELYSRSLSIGRLTRLAERVRLVDMAEQGASFLDLYRHLRERGAERRDAYFDAQQRVCRGGLV
EGGAPFTKDACYLAGLLEVYAFLAAVLRGGLRDEVLLVCGRIALDDIAVLAEALRAAGVLERP
5 RYLPGWLRWQTLTPYFAFTSFMDGIDLGPVERHFQELLRVAADARPAGEGRRRRGRPREG

15

20

10

SEQ ID No 30 (>ORF12)
MSESVAQLEEHRAALTGH CYRMLG SVVDADDAVQETMVRAWRS LDKFDGRSSLRTWLYRIATN
VCIDL RADRRARRARPIEEGPVGTVD DALETRPRTHWLEPV PDAHALPADIDAAERAMLRQSIR
LAFVAALQHLPPKQRAALLTEVLGWSAAEVADSLNTSVAAINSALQRRATLASRDLGDARP
SLPEPQSALLDRYVNAFERYD VDAL TALLHQDATLSMPPFTLWLRGHESIRAWLVGPGAGCRG
SRLIPTAASGSPAFAQYRPAPEGCHRAWALIVLDVAGDRIVSMTSFLDTETLFPFRFGLPLDLP
A

25

15

30

SEQ ID No 31 (>ORF13)
VTIASIDHRDQDLMTGPQAKAPARAAAPDAAPSRRVWAGRVLSGLATLFLTFDAAVKVLKLF
PAEASTAELGFP AHLVPTLGYLQIACLVAYLIPRTAVLGAILWTGYLGGAIAIHVRVENPLFS
HTLFPIYVAAFLWAGLWLRDRRVRALTAS PSSQGR

35

20

SEQ ID No 32 (>ORF14)
MTTKNPRKLFVNLSVRDLKRSMEFFSKLGFEFNPQFTDEKAACMVVSEEAYVMLLVESFFKTF
MKKEICSTSTHTEGLFALSCSSRAEVDDMVKKAVAAGGSHAMPQDHGFMYGWSFYDVGHHW
EVMWMDPKAIQP

40

25

SEQ ID No 33 (>ORF15)
MTPSERLDATFAALADPTRRAILARLASGEASVTELAKPFAMSQPAISKHLKVLERAGLISRG
RDAQRRPCRIEAKPLEDASGWL DNYRRFWEGSYERLDDLLEELKERESKGERSKR

45

50

55

5

100

SEQ ID No 34 (>ORF16)

10

VAPASAPAAGGRDAAPFLDEAAQWLRGEQAPASRPAGEGPAGRLPGRVLVADDNADMREYALR
LLVAEGWTVEAVADGRAALERARAHPPDLVLTDMMPRLDGFGLLRALRADDRTRGVAVVMLS
ARAGEEARVDSLEAGADDFLVKPFSAKELLARVRIHVELARRRREAEGQRQYLNDLFMQAPGP
5 IAILRGPEHVFEVFNPLYQRLVGGRSLVGEPIRAALPELEGQGIWELLDVAVRTGEPIVGKEL
15 PVRLDRRGDGTTEEVFFNFVYQPMRDRDGAVEGVFVFAFDVTDQVRARRRVEALVEALKLADQ
RKDEFLAMLAHELNRNPMASISLSLTLLDDADGDGPASARYREIARRQMGLVRLVDDLDDVSR
ITRGTVELRLEDVDLAAVVQSAAAVRPAVEARRHDVSLSVGPGDFGMRADATRLQVVTNLL
TNAKYTPPGGSISVRLTREAAVGAPEAVLRVDTGRGIPAAMLEKVFDFLTQVDQTIDRSTG
20 GLGLGLTLVRRLLLELHGGSVAAASAGPGQGSEFTVRLPLGPGAAPQAPSAGPPPPREGPPPA
10 QRDEPPPPPAQRAEPEAAADRRRLVVEDAEDVRRVMRAYIEALGHEVTVAVDGLEGVKKLL
ELRPEVAFVDIGLPGIDGYEVARRARAAPGGEALYLVALSGYGGPDDQARSRRAGFDLHLTKP
25 VVGATLQDVLTAAPT

25

15 9. DNA sequence according to claim 7 selected from the fol-
lowing

30

(a) open reading frames, and peptide sequences corresponding to
said open reading frames:

20

35

pEP0cos6_ORF1 sequences:

(1) nucleotide sequence

40

Seq ID No 35 (>pEP0cos6_ORF1.seq)

25 GGATCACCTGCGGCGCGATCGCCGACCTCGTGCTGGTGTTTCGGCTCGCTGGATGAGAAGCCGG
CGGCGCTACTGATAGAGACGGCGACGCCCGGGCTGCGGGTGGAGCGGTTGCGGGAGATGCTCG
GCTTTCGGGCGGCCCCACCTGGCGAAGCTGTCTTCGACGGTTGCGAGGTCCCCGAGGCTCAGC
45 TGATTGGCCGGCCCGGCTTTGCGCTGATGTATCTGGCCCCCTACGCCCTGGATTTGCGTTCGGG
TCAGCGTCGCCTGGGCCTGCCTGGGCATGATCCGCGCTTGCCTGGAGACCTGCGCACAGCACA
30 TCCTCACCCGCCGCACCTTCGGCCACCTGCTAGCCGATCACGGCATGATCCAAACCCTGATCA

50

55

5

10

CCAACCTGGGGATTACACCACCAGGCGACGCTGCTCCACACGCTGCAGGCCTGCCGCGCCAGGG
ATCGCGGCGACGTGACCGCCTCCGAGGCCACCCTCGCCGCCAAATACCTCGCGTCGCGGACGG
CGGTCCAGGAGACGACCAACGCGGTCCAGATCATGGGCGCGCTGGGCTGCGACGAGGAGGGCG
CGATCGCCCGCCACTTCCGCGACGCCAAGACGACCGAAATCATCGAAGGCAGCAACCAGATCA
5 TCGAGGCGCTGCTGGCCAAGAATCGCCCGCGCGGTGCGGACAACTATCGCCGCTTCTCG
15 ATGCGGAAGTCGAGCCCGGTGCGGCGGAGGCGCACCA

(2) peptide sequence

Seq ID No 36 (>pEP0cos6_ORF1.pep)

20

10 ITCGAIADLVLVFGSLDEKPAALLIETATPGLRVERLREMLGFRAAHLAKLSFDGCEVPEAQL
IGRPGFALMYLAPYALDFGRVSVAWACLG MIRACLETCAQHILTRRTFGHLLADHGM IQTLIT
NLGIHHQATLLHTLQACRARDRGDVTASEATLAAXYLASRTAVQETTNVQIMGALGCDEEGA
25 IARHFRDAKTTEIEGSNQIIEALLAKNIARAGRDNYRRFLDAEVEPGRAGGAP*

25

15 pEP0cos6_ORF2 sequences:

(1) nucleotide sequence

Seq ID No 37 (>pEP0cos6_ORF2.seq)

30

ATGACGAGCGCGGTCCCGACCGTCAAACCAGCCTGCTCGACGACTTCGAGCGCGTCGCCGAC
20 GTCGATCCAGAGCGGATCGCCGTCCACGCGAGCGAGACGAGCCTGCGCTATGGCGACATGAAT
35 GCGCGCGCCAACCGCATTGCCCCACGGGCTACGGGCGCGCGGGATCGGGCCCAATCAAATCGTG
GCGGTGGCGATGGCCCGCACGCCCGAGCTGATGATCGTGCTGTACGGCATCCTCAAGGCCGGC
GCGGCCTACATGCCCATCGCCCGGACGCGCCCGCTGCGCCGCGATCATATGCTGCGCGAG
40 AGCCAGGCTGCTCTGATGATCGCCGACGAAGAGATCGCGGGACTCGCGGCCCGGTGCTGACG
25 CCGGCCGACCCGTTCTTCGCGGCCATGCCGACCACAACCCGAGCCGCGTCACGACCCGACC
GACCTGATTTACGTCATCTACACCTCGGGCTCGACCGGCCAGCCCAAGGGCGTGCCCATGGAG
CACCGCGCCGTGTGGAATCGCTGACTTGATGCAGGCCCAGTATCCAATCGACACGCAGGAC
45 GTGATCCTCCAAAAGACGCCGATCGTCTTCGACGTGTCGGTCTGGGAGCTGTTCTGGTGGCCG
CTGGCCGGCGCCTCGGTGGCCCTGCTGCCGCAATCCATGGAGAAGTTCCCCCTGGGCGATATCG
30 GCGACGGTGGCGCGGTGCGGGGTGACGGTGATGCATTTCTGATCATCGATGCTGATGGCCTTC

50

55

5

10

15

20

25

CTTCAGGTGGTGGCGGGCCGGCCCCGAGATGGCGGACCAGATGAAGGGCCTGCGCTACGTCTTC
TGCAGCGGCGAGGCCCTGGCGCCGGCCCCACGTGTTCAGCCTTTCAGGAGCACATCAACCGAGCG
GGCAGCATCAGCTTGACCAACCTCTATGGACCCACCGAGGCGGCGGTTCGACGTACGTACTTC
GACTGCCCCCGGCGCGTCACTCGCGCGGTGCCGATCGGACGAGCGATCACCGGCATCCAG
5 CTGCTGGTCATGCGCGACGGCGTGCCTCAGCCGCCCGCGTTCGAGGGTGAGCTCGCCATCGGC
GGCGTTGGTTTGGCGCGCGGTACATCTCACGGCCAGACCTGACCGCCGACCGGTTCTGTCCG
15 CATCCAGGCGGCGACGGCCAGCGGCTCTACCGCACCGGCGATCTGGTGCGCAGGGACGCGGAC
GGCGAGCTGGTCTTCTGGGGCGCATCGACCATCAGGTGAAAATTCGCGGTCTGCGCATCGAG
CCCCGGGAAATCGAGGCCAGATCAGCGCCCATCCCGATGTGGCCGACTGCGCGCTGATTATC
20 GAGCAGGACTCGGAAACCTGCCCLAGCTGACCGCCTACATTGTCTGTGGCGCGACCGGGCTTG
10 ACCCGGAAGGCGCTGCTACAGTTCCTGGGCGCGCGGTGCCCGACTACATGCTCCCGAACCGC
TTCCTGACCTCTACGGAGCTGCCCCGTGACCGCCAACGGTAAGCGCGACTGGCGCGCGCTGCTC
GGCCCCGCTCGAGACCCTGCCTCTCCCTTTCTCC

15 (2) peptide sequence

Seq ID No 38 (>pEPOcosf_ORF2.pep)

30

35

40

MTSAVPTRQTSLLDDFERVADVDPERIAVHASETSLRYGDMNARANRIAHGLRARGIGPNQIV
AVAMARTPELMIVLYGILKAGAAYMPIARDAPPLRRDHMLRESQAALMIADDEIAGLAARVLT
PADPFFAAMPDHNPEPRHDPTDLIYVIYTSGSTGQPKGVMESHRAVWNRLTWMQAQYPIDTQD
20 VILQKTPIVFDVSVWELFWWPLAGASVALLPQSMEKFPWAISATVARCGVTVMHFVPSMLMAF
LQVVAGRPEMADQMKGLRYVFCSGEALAPAHVSAFQEHINRAGSISLTNLYGPTEAAVDVSYF
DCPPGASLARVPIGRAITGIQLLVMRDGVPOPPGVEGELAIGGVGLARGYISRPDLTADRFVP
HPGGDGQRLYRTGDLVRRDADGELVFLGRIDHQVKIRGLRIEPEIEAQISAHPDVADCALII
40 EQDSETLPKLTAYIVVARPGLTRKALLQFLGARLPDYMLPNRFLTLTELPVTANGKRDWRALL
25 GPLETLPLPFS

45

50

55

5

103

pEPOcos6_ORF3 sequences:

10

(1) nucleotide sequence

Seq ID No 39 (>pEPOcos6_ORF3.seq)

15

20

25

30

35

5 ATGTTACACCCGATTCCCACCGACCGTTTCGCCCTGAGCCGACCGCTCTTTCGCGGGTACCTC
GCGCACGATCCGATCGTGCGAGGGCGTGCTGGCGGGCGACCATCCAGGCTGGGTCTTGGTGGAC
CGCGAGCCCCGAGCCGCGCACGGCGCTGCTGTGGGCCTTTTCCGATCGGCTCTTCTGCGTGGGC
GCAGCTGACACGCTGACCCCGCACGCGCTGGCCGAGCTGTTCCACGACCGACTGATCCCCCAG
GCCCGTAAGATCGGGCAGCCGTTTTTCCAGGTTTCAGGGCGAGACGGTCGACACCTGGTTCGGAC
10 CACCTGCATCAGGTGTCGCCGCACCGACAGTCTCCTTCCGCCAGGCATTCCGCTTCGACCGC
GACCTCTTCGAGCGGCTGCCAACCAAGCCGGAGCTGGCAGAGGCGCGGCTCGTGCCAATCGAC
GCGCGGCTGCTGGCCGAACAGGCTGATCTGCGCGAGCGGATACTGGCCTCCTGGTCCAGCGAA
GCTGCCTTCCATGCGCGCGGTTTTCGGCTTCTGCTACCGCGTAGGTGACCAGCTGCCGAGCGTG
15 TGCCTGGCATCGCACGTAGGCGGCGGCGCGGCCGAGCTGAGCATCAACACCGAGCTCGAAGCG
CGCAATCGAGGTATGGCAACGCGGCTGTGCCGCGGTTTTTCATCGCCGAATCGCTGCAGCGCGGC
CTGACGCCTTGCTGGGGCACCAGACCTTTTCGCTGCCGTCAATCGCGCTGGCCCAAGAAGCTC
GGTTTTTCATCCCGACCTTCACCTTCCCCACCTACTGCTTCGCGACCGGCACCGAACAGCCGGAC
30 GACAACTTCCTAGGCGAGCTGTACTACAGGAATCGCGCATCGCCGGAAGTGGGACCGATGAG
CCGCAAGCGGTTTCGGCTGGCGCGGGGTTGGAGCCTGGCCGCGACACCGAGCGTGCCGCGAGC
20 TTCGCCGCACGCGCCCTGGCCGAAAGGTGGGCGGCCACTCGACTCTGGCCACCGATCCGGAT
TTCGCCCCGATTGCGCGCCAGCGCCGCTGGCCCCGCTCAATGTCCCT

(2) peptide sequence

40

Seq ID No 40 (>pEPOcos6_ORF3.pep)

45

50

25 MLHPIPTDRFALSRLFRGYLAHDPIVQVLAGDHPGWVLVDREPEPRTALLWAFSDRLFVVG
AADTLTPHALAELFHDRLIPOARKIQPFQVQGETVDTWSDHLHQVSPHATVSVFRQAFRFDR
DLFERLPKPELAEARLVPIDARLLAEQADLRERILASWSSEAAFHARGFGFCYRVGDQLPSV
45 CLASHVGGGAAELSIINTELEARNRGMATRLCRRFIAESLQRGLTPCWGTETFRLPISIALAQKL
GFIPTFTFPTYCFATGTEQPDDNFLGELYRESRIAGSGTDEPQAVRLARGWSLAGDTERAAS
30 FAARALAEGWAGHSTLATDPDFAPLRASAAWPRLNVP

55

5

104

10

pEPOcos6_ORF4 sequences:**(1) nucleotide sequence**

5 Seq ID No 41 (>pEPOcos6_ORF4.seq)

15

ATGATTTGTCACTCCACCGCTTCATTTTCTCCACGTTCCCAAGGTCGCCGGCACAAGCGTC
AAGGACGTCTCGGCAAGAGCTATTCCAGGAGGACCAGGTCACGTTCCAGATCGCTCCCAAT
CCCCACTACCCACCTGAATGGACTGCGCCTTACGAGGAGCACATTATTGCCGCTGAATTGAAG
AGCCAGTTGGCGCCGAAATTTGGGACGATTACTTCAAGTTCGCCTTCGTGCGCCATCCGCTC
10 GACTGGGCGGTCTCCAATTACTTCTTCTTCTGCGCGACCGCAAAGGCCATCCGGCCCACGAA
TTCCTGGAGCGGAAGGGCTTCGCCGGTACCATGGACATGTTTTTCGGAGCGGCCGGGCGCCAT
CCGCTGGTCGCCGGCATGCGCTTCAGCCAATGGGAGTTCTTGTGCGACAGCGAGGCGCGGACG
25 CTGGTGGACTTCGTTGGCAAGTACGAGCGGCTCGAGCAGGACTTCGCCCGCGTGTGTATCCGC
ATCGGGCTGACCCCGCCGACTTGCCGTGCCTCAACCAGACTCGCCACCAATCCTTTACCACT
15 TACTACGACGAGGCTTTGATGCGCCAAGTCAGCCGCGCGTTAGCTCGCGATTTGAAATTTTT
GATTATGCC

30

(2) peptide sequence

Seq ID No 42 (>pEPOcos6_ORF4.pep)

35

20 MICHSHRFIFLHVPKVAGTSVKDVLGQELFQEDQVTFQIAPNPHYPPPEWTAPYEEHIIAAELK
SQLAPEIWDDYFKFAFVRHPLDWAVSNYFFFLRDRKGHPAHEFLERKGFAGTMDMFFGAAGRH
PLVAGMRFSQWEFLCDSEGRTLVDFVGKYERLEQDFAAVCIRIGLTPDLPLCLNQTRHQSFTS
YYDEALMRQVSRALARDFEIFDYA

40

25 **pEPOcos6_ORF5 sequences:****(1) nucleotide sequence**

45

Seq ID No 43 (>pEPOcos6_ORF5.seq)

ATGAAAGTGGACAAGCGGAATGTCGACGACATTCTCGGACTCACTCCGACACAGACAGGCATC
30 TTGTACCACTACCTGCTGGACCCGAGGCCGACGCCTATTTGAAACAATTGACGCTGCACCTG

50

55

5

105

10

15

20

25

30

35

40

45

50

55

GAGGGGCGCTCGACGTAGCGCGCTTCCGCCGCGCTGGGAGCGCGTGGTGGCGGCTCACGAC
CAGCTGCGCGCCGTGTTTCGCTGGCAAGGGATCGAACACCCGGTGAGATCATCCTCAAGCAG
CACGTGCCGGACCTGGAGTTGGCGGAGGTCCCGCGCGACGCCGATCCGGCAGCCTTCCTGGCG
CAATGGGTGCGGGCCGACCGGGCGCGCAAGTTCGACTTCGAGACGGTGCCTTTTCGCATCGGC
5 CTCTGCCGACTGATACCAACATCACGTGATGCTGCTCAGCAATCACCATATCCTGATGGAC
GGTTGGAGTACGGGCTGATTCTGCGGGACTTCCTCGCTGCTACGGCGACTCCGAAAAGTGG
CGGCCACGCACCCGAACGCACTTCAAGGCGTTCATCAAGTGGCACCAGAACCGGCCACGCCGG
GGCGAGGAGCGATTTTGGCGCGACCTGTTGCGCGATGCGCCGACGGCGGCTTTCCCGCGCTG
GGCGTCGAAGAAGGCACCCGCCACTCGCTTGACTTCGGCGCCCGCAGCCGCGCTCTCGACGAC
10 CGCTTGACCCAAGGCTTGCGCGACATGGCTCGCGACCTCGACGTCACCTCGCCGCGATGCTC
CATACCGCTTGGGGCTTCTACTCCAGCGCTACCAGAACAGCTGCGAAGTGATATTCGGGACC
ACCGTTTCCGGCCGCAACGTGAGCTCGCCGGCCTCGACGAGGTGGTGGCTTGTTCATCAAC
ACGATTCCGTTCCGCTTCTCGGCCGCGCGCGACGACGCGCCGTGAGGCCTTCGTTGCGGTA
CAGCGCAATCTGCTGGCGAGAAGCGAGTTGAAGCCACCCCGCTGGTGGACATCAAGGGCTGG
15 AGTGGTCTCGGTCCGGGCGCGGAAGTGTTCGACACCATCCTGGTTCATCGAGAACTATCCCTTG
GACCGCGCTATCTTCGAGAGTGATTCCAGCCTGCGGTTGACCGACCACCAAATCTTCGAGCGC
ACCAATTACGGGCTGACCTGACCATCGAGACCTTCAGCCGGTTGCACGTGACGTAGCCCAT
CGCCGTGACCTGCTGGGCGACGCGGCCGCTGAGCGAATGCTAGATCATTTACCGGCCTGCTC
CAAGCCATGCTGCGCTTCCCTCACCAGCCGTTGCGCGCCTCGAGATGAAAAGCGAACACGAG
20 GCCCACCGCTCCTGCACCAACTCAACCAACGCGTCAGCCGCTGCCGTCCCAATCGGCTTTC
CACCAGTTGTTCTTCGAGCAGGCCCAGGCCGATGGGGCACGACCGCGCTGTGGTGGCGCGCC
ACGCGCTGGACCTACGGCCAGCTGCTGGAACGTGCCCTGCGTCTGGCGGGACGGCTGCAGGAA
GCCGGCTTCGCCCCGAGGCGATGTGCGCCCGCTCAGCCTCGGCCCGGTTCCGGATCTGATTCCC
GGTTTGCTGGGCCGCTGTTTCGCCGCGCGCGCCTACCTGCCGCTCGATCCACCCCTGCCGGCC
25 CAGCGCTCGCGGTTTCATCCTCGACGATGCCGTTGCCGCTTCCTGATCAGCGACGCGCCACTC
GCGGGGCCCACGCCGATCCATCCGACCCTGCCGGCGCCAGCCCCGTTGACGTCAATTTTGCC
TGTCAGGACGGCGCCGCGCAGCCCGCTACCTGATCTACACCTCGGGCTCCACCGGCCAGCCC
AAAGGCGTCTGGGTTAGCCACCGCAACCTGATCAACTTCCTGACGGGCATGAGCGCAATCCTG
CCGCTCGCGGCCGACGACGTGTTCTCTCGCTGACTACCGTGTGTTTCGACATTTTCGGGCTC
30 GAGACGTGGTTCCCGCTCAGCCCGGGCTGCACGATCGTCTTGGGCACGCGCGCCGAGCAGTTG

5

10

15

20

25

30

35

40

45

50

55

GACCCGGCCGCGGCTGCCAAGGCCATCTCCTGCCATGGCGTCACGTTTACCAGGCGACGCCA
TCGCGACTCCAAC'TTCAACTGGAGCACCCACATTTGTCCGCGCCATCGGCTCCCTGACGACC
CTGCTGGTAGGCGGCGAACCCTCCCAGCCGAGCTGCTGCGGCGCGTACGCGAAGTGACCGAT
GCGCGTATCTTCAACCTCTACGCTCCCACCGAAACCACCATCTGGTCCACAGCCGGGGAGGTC
5 ACCGCGGCGGACGTCCCGGATATCGGCCGCCCGATCGCAAATACCGGCGTTTTCCTTCTGGCG
CGAGACGGCTCGATCCAGCCCGCGGCTGGTGGGCGAGTTGTGCATCGCCGGCGAGGGCGTG
GCGTTGGGCTACCACCGACGGCCGACCTGAACCGAGAACGGTTTCGCGAGATTCCGCCGGGC
CGCCTGCCCTTTGCCGGCAAGCTCTACCACACCGGCGACCTGGCCCGCTGGACCGAAGACGGA
CGGCTCCTCTGCCTGGGCCGTCTGGACGACCAGCTCAAAGTGCGCGGCCATCGCGTCGAGCCG
10 GCGGAGATCGAGGCAGTGATGGCGCGCCACCGGCGGTACGCAGGCGGTGGTCGTACCCGG
CCGCGCAACGGCGAGCCGGTCTTGCTCGGGTTCTGGACTGCGGAAGGTGAGCCGATGCCAGAG
GAAGCGCTGAGCGCTTACCTGGCCGACCGACTGCCGAGCTACATGGTACCCGAACGGTGATC
CTCATGAAGGCCATGCCGCTAACCGGCAACGGCAAGATCGACCGGCGCGCCCTACCCAATCCC
TTGCGCTTGACCGAGTCGACCCGGCAGGCGGCGCGCACCTTGGCCCGCACCGCCGGCGAG
15 CATCGGGTTGCCGAGCTGTGGCAGGCCCTTGTTGCGACGCGAGGCGATCGGCTTGGACGAACCC
TTTTTTTAGGCCGGCGGAACTCATTCGGCTTGATTTCGGCTTACGCCAAGCTGGAATCCGCC
TTCCGGAAGTCGTTCCCGATCACCGATTGTGTTCCAGCATAACAGTATTTCGAGCCAGGCAGAA
ATGCTGAGCGGCTCGTCCGTCGAGGCGCGCTCGCGGAGCCGTGCCGCAACCCCGGCGGCC
GCCGCCAAGTTGCCTCCTCGGCAGCTAAATCCCCAGGGGAGCGCGGCGGCGAGCGACGTCG
20 AGCGGCCTGACCGCGCAACCGCCCCAACCCCACTTCCGGCCCATCGCCGTTATCGGCCTCGCC
GGCCGATTCCCCGCCGACCCGACCTCGACGCCTTCCTTGAAGTGTCTACGGAGGTCGCTGC
GGCATTCGCTTCTTCAGCCAAGCCGAGCTGCGCGACGAGGGTCTCGACGCGAATCGAATCGCG
TGTCATAACTATGTCCCGGCCAAAGGTTTCCTCGACCGGGCCGACCACTTTGATGCCGACTTC
25 TTCGGCATCCCGCCGCGGACGCGAGAAATCACCGATCCGCAAATTCGGCTTCTGCTTGAGTGC
TGCTGGAACGCGCTGGAGCATGCCGCTACCCGCCCGGCGGCGGAGATCGGGCTCTTCGCC
GGCTCCTCGGCCAACTATCACTGGCTCGAATACGTGGGCATTTCCGAGGAGAGCAGCAATCGA
TTCGCCGTATGATTCAAAACGAAAAGGACTACCTGGCCACGCGGATCGCCTACCAGCTCGAT
40 TTGAAGGGCATTGCCGTACCGTGCAAACGGCCTGCTCGTCGTCGCTGACCGCGGTGAGCTG
GCCTGCGATGCGTTACACGCCGGCCGCGTGACCATGGCTTTGGCTGGTGGCGTTGGTCTGACC
30 TATCCGTTGCGCGCCGATACCTGCACGAGGATGGAATGATCTTCTCCCCGACGGTCCGGTGC

5

107

10

5

15

20

10

25

15

30

20

35

40

25

45

30

50

55

CGGGCCTTCGACGCCCAGGCGGCCGGCACGGTCTGCGGCAACGGTCTGGGCATGGTGGTGCTG
AAACAGCTCGACGCGGCGCTGGCCGACGGCGATGCCATCCACGCTCTGATTAAGGGCATCGCG
GCCAACAAACGACGCGCGGCCAAGATCGGCTACACGGCGCCCTCGCAGAACGGTCAGGCGCGG
GTGATCCGCGCCGCCCATAGGCTCGCCCAAGTCGCGCCGAGACCATCGGCTATGTAGAAGCC
CACGGTTCGGGACGCGCGCTGGGCGATCCGATCGAGGTGGCGGGCCTGACCGAGGCCTTTGAC
AGCCCCGCTCGCGGCTTCTGCGCCTTGGGTTCGGTCAAGTCGAATGTGGGTCAATTCGATGCG
GCAGCGGGCATCGCGGTTTCATCAAGGCGGTGCTCTCGCTGTCCCATCGGACCCTGTTTCGCC
AGCCTCCACGTGACACGCCCCAACCCGAGATCCCGTTCGCCGACGGTCCGTTCAGGTCAAC
ACGGAGACCCGGCCCTGGCCAGCTGCCGACCATCCCGCCGCGCCGGCGTCAGTCTTTCGGC
ATCGGCGGCACCAACGTGCACGCCGTCTTGAAGAGGCGCCGAGTTGGCCGAGCACGCGGGG
CGGCGGCGCGAGCGGCAGCTGTTCTGGTCTCGGCGCGGACTGCAGCCGATCTGGAGCGACGC
ACCGCGGCGCTGGTCCGCCACCTGGCCGCGCATCCGGACCTCGCACCAGATGACGTTGCCTTT
ACCTTGACGCGGGCCGCAAACCGATGACCCACCGTCGTTTCTGGTCCGCCCGACCTCGCG
GAAGCCGCGCGCGTCTGGCCGAGCCCGATCCAGTCAAATCCGCCGCGGCGCGCCGACCGC
TGCCAGGTCTGGATGTTCCGCCGTCTCGGCTCTCAATACCCCGGCATGTGTGGCGGCCTCTAT
CGCACCGAGCCGGCCTTTCGCGAGCAAGTCGACCGCTGTTTCGACCTCCTCGCGCCGCGTTGC
GATTTGAAGCCCTCGCTCTTCCCGAGCCCGATCAGGCCATCGACGCATCAGCCCTCGCGGCC
ATCGACACCGCCAGATCGCGTCTTTCGTCTGCGAATACGCGCTCGCACGGATGCTGGAAGGC
TGGGGGCTGCGTCCGATCGGCTGATCGGTTACAGTTTCGGCGAATACGTGGCCGCCTGCCTG
GCCGGCGTCTTCTCCCTGCCCCGACGCCTTGGCAATCGTCCGCGAGCGTGGCCGATCCTGGCG
GCGGCCGAGCCGGGCGCGATGGTCAGCGTGCCCCCTTCCGGCCGAGCGCGTCGCGTCTGCTG
GAGCCGCCGCTTGCTTGGCCATTGACAACGGCCCCCTCATGCGTGGTGTCCGGGCCGGTCGAA
CCGGTGCGCACCTTCACCGCTCGCATGAAGCCGGACCGGGTCTGGGTGACGCCGCTCCAGGCC
GAGCGCCCGATGCATTGCGCGCTGATGGCCGAGGCCGGCGGCTCACTGCGCGCCATGTTGGCC
GGGTTCCGCCTGAATGCGCCGCGAATCCCGATCTTAAGCAATGTTACAGGAACCTACCTAACC
GACGAGCAGGCCCGAGACCCGATTACTGGGCCCGTCACCTGTGCGGCAACGTTTCGCTTCGCC
GACGGTGTGCGAACCTTGTGGCCGAGCGCGATCCGGTGTTCCTTGAATTCGGGCGCGGCCGC
GATCTGAGCTCCTTGGTGCGCCACCAGATGCCGAAGGCGCGACGAGCCGATCGCACTGATC
CGTCATCGCGAAGATCCGGTGCGCGACGAAGACCTCCTGCTCGATGGCTTGGGCCGCTGCTTC
CTGCGTGGGGCGACCCTCCACGGGCAGGCCTTGTACGCCGGCCGAGGCTGCCGCCGCGTGCCG

5

108

10

15

20

25

30

35

40

45

50

55

CTGCCCGGTTACCGGTTCCAGGGTCCACGCTGCATGCCGGCCCCGCGCCGGACTGCCCGGCCTG
GCGCGACCGACCGTGGGAGCGACCAACCATCAGCTACCGACAGCCTGGAAGCGGCGCCGCGC
TTGGCGGCTGTGAATCGCTCGCGCCGCAATCCTGGTTGGTATTCAGCGACGGCAGCGAATTG
GCGGGCGAGCTGGTGGCCGGCCTGCGCGCTTCCGGTTGCGCGACCAACCTCGTGAAGGTGGG
5 CTGGCGTTTCGCGCGCTTCGCGGGCGGCTTCCGCGCGAATCCCCGCGAGGAACAAGATCTCGCA
15 CAGCTGTTTCGCGACCTGTTCGGCCGAAGCGATGCTGCCCCACCCACATCCTGCACCTGCTCAGC
CTGCCGTCGCCGGAGCGCGACTCGCCGCTGGCGCGCCTGGAGCACCTCACCGAGCTGGGCTTC
CACCATCTGCTGGCCCTGGCCCGCAACTGGAGGCGGTGCGCGCCCCGAGGTCCGCCTCGCC
GTGGTGACAACCGGCCCTGGCGGCGATTGGCGGCGAGTCCGAGCTGCGGCCCCGAGGTTCGGGCTG
20 TTGCGGGGACCTGTCCGCGTGATTCCCTTTGAATTCCCGAACTTTCGGCTGCGCCTGATCGAC
10 CTCGACTCGGCCGATCCCATCTGGCGTAGCGGTTGTGAGCCGTTGCTGCGCGAAATGGGCGCT
GCCCCGGGACCTGAAGAAATCGCGCTGCGCGGCACACGCGTTGGGAGTTGGGCTACGAGCCG
GTCGAGGGGGGCACCGTGAGCACCATCTCCTCGCGACTGCGCGAGGGCGGCGTCTATCTGATC
25 ACCGTTGGCCTCGGCGGCCTGGGTCTGGCCTTGGCCCCGTACCTCGCCCCGAAGTACCGCGCC
15 ACCCTGATCCTCGCTGGCCGGCGAGGCGCGCCGGCGCGGAGCTCTGGCACAGGCGCCAGCG
GAGTTCTGACCGGTTCGAGCTGCGATCGCACAGATGGAGGAGTGTGGCGCCCGCGTGATTCCC
30 GTCGCGCTCGACGTACCGACGCCGACCAAGTGAACGCGTTGTTTCGCCACCATAGAAGCTACG
GTCGGCAAGATTGAAGGCGTTTTCCACATGGCTGGCATCGTTGACGGCGGCATCATTCGAACG
CGCACGCGCGCTGCCAGCGACGCCGTGCTGGCGCCCCAAAACGGTCGGAACCTGGATTCTCGAT
20 CGGGCTCTCCGCGGCGCCGGTGGCGGCTTCTGGTGCTGTAATCCTCGATCAACGCGGTCTGTC
35 GCGCCCTTCGGCCAGGTTGCCTACGCCGCCGCAACGCCTTCTCTGACGCCTTCGCCAGCGCC
CACGAACACGACGAGCGTCTTTTCCGCGTCAGCATCGGTTGGGACACCTGGCGCGAGGCCGGC
ATGGCCGTCGATGCCGCCCCGCGCCGCGCGACCAAGGCCCGCTCGAAGGGCTTAGCGACGAG
40 CAGGGCTTGGCGCTGCTCGAAAGCGCCTTGGTTCGGTTGCGAACCAGCGACTCCTCGTCTCCATC
25 AGCGAACTGCGCGCTCGACTAGCCGAGCATCATCGCAACGGCGGCATTCCCCGGTTGCTCGGG
CCCCGCGCCAACGAGGCGGGTGCAGCTGATTCCGGCGAGGAGGGCGCCACGCAAGACGCGTCG
CCGGCCCGTCGCGCCCGTCCCAGTCTGGTTCGTGGCCTTCGCGCCGGCCGCAACGAGCTGGAG
45 CGCCGATCGTGGCCATCATCGCGCCCTACCTGCGGCTCGGTTCAGGTGGGCGTCGACGACAAC
TTCAACGATTTGGGCGCCACCTCGCTCGACCTCATCCAGATCGCCCAACGCCTCGGTTCGCGAG
30 TTGGGCCGCGATGTCCCTGTGCTCTCGCTCTACCAACACCGCACCGTACGCGGGCTGAGCCGC

5
10
TTCCTCGGCGGCGCGCTCCAATCCGCGCGGTCCGGCGTCCCGACGGGCGCTGCCGCACCGGGC
GCGGCCACGCGGGGGTTGCCACCCCGCGCGGCCACAACCGTCGCGCCAGCACCTGGAAAAA
CGCCGTCAATTGAGGAAAAAGGGGGGCCTTCCCATCATGAG

5 (2) peptide sequence

15 Seq ID No 44 (>pEPOcos6_ORF5.pep)

10
20
10
25
15
30
35
40
25
45
30
50
55
MKVDKRNVDILGLTPTQTGILYHYLLDPQADAYFEQLTLHLEGPLDVARFRRWERVVAHD
QLRAVFRWQGIIEHPVQIILKQHPDLELAEVPRDADPAFLAQWVAADRARKFDFETVPFRIG
LCRTDTQHHVMLLSNHHILMDGWSTGLILRDFLACYGDSENWRPRTRTHFKAFIKWHQNRPRR
10 GEERFWRDLRLDAPDGGFPRLGVEEGTRHSLDFGARSRALDDRLTQGLRDMARDLDVTLAAML
HTAWGLLLQRYQNSCEVIFGTTVSGRNVELAGLDEVVGLFINTIPFRFSAAAATTPEAFRAV
QRNLLARSEFEATPLVDIKGWSGLGPGAELFDITLVIENYPLDRAIFESDSSLRLTDHQIFER
25 TNYGLTLTIETFSRLHVTLAHRRDLLGDAAERMLDHFTGLLQAMLRFPHPQPFARLEMKSEHE
AHRVLHQINQTRQPLPSQSAFHQLFFEQAQADGARPALWCGATRWTYQQLLERALRLAGRLQE
15 AGFARGDVAAVSLGPVPDLIPGLLGPLFAGGAYLPDPTLPAQRSRIFILDDAGCRFLISDAPL
AGPTPIHPDPAGASPVVDVIFACQDGAAPAYLIYTSGSTGQPKGVMVSHRNLINFLTGMSSAIL
30 PVAADDVFLSLTTSVFDIFGLETWFLSRGCTIVLGTAEQLDPAKAAKAI SCHGVTVYQATP
SRLQLQLEHPTFVRAIGSLTTLVGGEPPLAELLRRVREVTDAIFNLVYPTETTITWSTAGEV
TAADVDPDGRPIANTGVFLLARDGSIQPPGLVGELCIAGEGVALGYHRRPDLNRERFREIPPG
20 RLFFACKLYHTGDLARWTEDEGRLLCLGRLDDQLKVRGHRVEPGEIEAVMARHPAVTQAVVTR
PRNGEPVLVGFWTAEGEPMPEEALSAYLADRLPSYMPERCI LMKAMPLTGNGKIDRRALPNP
35 FALTESTRQAAPRTLARTAGEHRVAELWQALLRREAIGLDEFFQAGGNSFGLIRLHAKLESA
FGKSFPITDLFQHTSIRSQAEMLSGSSVEAPLAGAVPQPPAAAAQVASSAAKSPGERGAAATS
40 SGLTAQPPQPHFRPIAVIGLAGRFPAAPDLDAFLELLTEGRCCGIRFFSQAE LRDEGLDANRIA
25 CHYVPAKGFLDRADHFDADFFGI PPRDAEITDPQIRLLLECCWNALEHAGYPPGGGEIGLFA
GSSANYHWLEYVGI SEESSNRFVMIQNEKDYLATRIAYQLDLKGI AVTVQTACSSSLTAVEL
ACDALHAGRVTMALAGGVGLTYPLRAGYLHEDGMI FSPDGRCAFDAQAAGTVCCNGLGMVVL
45 KQLDAALADGD A I HAVIKGIAANN DGA AKIGYTAPSQNGQARVIRAAHRLAQVAPETIGYVEA
HGSSTPLGDCPIEVAGLTEAFDS PRRGFCALGSVKS NVGHLDAAAGIAGFIKAVLSLSHRTLFA
30 SLFVDTPNPQIPFADGPFQVNTETRPWPAADHPRRAGVSSFGIGCTNVHAVLEEAPQLAEHAG

5

110

10

5

15

20

10

25

15

30

20 pEP0cos6_ORF6 sequences:

35

(1) nucleotide sequence

Seq ID No 45 (>pEP0cos6_ORF6.seq)

40

25

45

30

50

55

RRRERQLFLVSARTAADLERRTAALVRHLAAHPDLAPDDVAFTLHAGRKPMTHRRFLVAADLA
EAAARLAEPDPVKSAARADRCQVWMFAGLGSQYPGMCGLYRTEPAFREQVDRCFDLLAPRC
DLKPSLFPPEPDQAIDASALAAIDTAQIAVFVCEYALARMLEGWGLRPDRLIGYSFGEYVAACL
AGVFSLPDALAIVRERGRILAAEPGAMVSVPLPAERVASLLEPPLALAIDNGPSCVVS GPVE
PVRTFTARMKRDRVWVTPLQAERPMHSPLMAEAGGSLRAMLAGFRLNAPRIPILSNVTGTYLT
DEQARDPDYWARHLCCNVRFADGVRTLLAERDPVFLEFGPGRDLSSSLVRHQMPGADepIALI
RHREDPVRDEDLDDGLGRCFLRGATLHGQALYAGRGCRVPLPGYPFQGPRCMPARAGLPGL
ARPTVGATTISYRPAWKRAPRLAAVESLAPQSWLVFSDGSELAGELVAGLRASGCATTLVEGG
LAFARFAGGFRANPREEQDLAQLFATLSAEAMLPTHILHLLSLSPERDSPLARLEHLTELGF
HHLLALARQLEAVGAPEVRLAVVTGLAAIGGESELRPEVGLLRGPVRVIPFEFPNLRRLRID
LDSADPIWRSGCEPLLREMGAPGPPEEIALRGTSRWELGYEPVEGGTVSTISSRLREGGVYLI
TGGLGGLGLALARHLARKYRATLILAGRRGAPARELWHQAPAEFVPVAAAIAQMEECGARVIP
VALDVTADQVNALFATIEATVGKIEGVFHMAGIVDGGIIRTRRAASDAVLAPKTVGTWILD
RALRGAGGRFLVLYSSINAVVAPFGQVAYAAANAFLDAFASAHEHDERLFRVSGWDTWREAG
MAVDAARARGDQAPLEGLSDEQGLRLLESALVGCEPRLLVSISELRARLAEHHRNGGIPRLIG
PRANEAGAADSGE3GATQDASPARRARPDLVAFAPAGNELERRIVAIIGAYLRLGQVGVDN
FNDLGATSLDLIQIAQRLGRELGRDVPVVSlyQHRTVRGLSRFLGALQSARSGVPTGAAAPG
AATPGVATPPRPQPSRQHLEKRRQLRKKGGPSHHE

ATGAGTGAAGTATCCATTGCCCCGGCTTGGACATCGCGGTCATCGGCATGGCCTGCCGCTTT
CCCCGGTCCCCGCAACCTCGCCGAGTATTGGGCCAACCTGATCGAAGGCCTCGAAACGCTCAGC
TTCTTCAGCGAAGAGGAGCTGCGTGAGGCCGGCTGCGATCCGGTCCAACGGCCAGCACAAAC
TACGTGCGCACCAAGGGCCTGCTCCCTGACGCAGACCGTTTCGACGCCGATTTTTTTGGTTAT
TCCCCGCGCGAAGCCCAGGTGATGGACCCCCAGATCCGCGTCTTCACGAGGTCTGTTGGCAG
GCGCTGGAGCACGCGGGCTACAACCCGCATCGCCACACCGGCACGATCGGCCTGTTGCGCCGC
GCGCGCCCAACGTTTTTTGGGAGTTTCTCTCTATCGGTCCGATGCCGCCAATTTAGGCAAC

5

10

5

15

20

10

25

15

30

35

40

25

45

30

50

55

TTCACGCTGGGCCTGCACAACAACAAGGACTACCTGAGCTCGCGCATCGCCTACAACCTCAAC
CTGACAGGGCCCAGCTACACCCTGTTACCGCCTGCTCGACCTCGATGGTCGCCATCCACCAG
GCCGCTCCAGGCGCTGCTCAACGGCGAATGCGACCTGTGCATGGCCGGCTCGGTCTCCATTACG
CTGCCACTGGTTGCCGGCTACACCTACACGCCGGCATGATCGTCTCGCCCGACGGCCATTGC
CGCACCTTCGACGCAGGCGCCAATGGCACTGTCTACGGCGACGGGGCCGGCGTGGTCTGTTCTC
AAGCGGGCCGAGGATGCGTTGGCCGACGGCGACCACATATTTGCGCTCATCAAGGGCTCGGCG
CTCAACAACGATGGCAGTCGCAAGACCGGCTACACCGCGCCAGCGTCAGGGGCGAGGTGGAG
GTGATCCGCGCGGCGATGAACCTGGCGGAGGTGAGCCGAGCGGATCAGCTACGTGGAACC
CACGGGACGGGCACCACGGTGGGCGATCCGCTGGAGTTCAGGGCGCTAAAGGAGGCCTTCGGA
GGTGGCTGCAAGGCCTTCTGTGGATTGGGTTGCGTCAAGCCGAACATCGCCCATCTGGACGTG
ACGTCGGGGATCGCGAGCTTCATCAAGCTGGTCTTGGCGCTGGAGCACCGCATCCTACCGCCC
ACGCTCCACTTCCAACCTGCCCAACCCGAAGATGGATGTGGTCGATAGCCCCCTTCTACATCGTG
GCTGAGCGCGAACCTTGGCGGAAGATCTGCTGCCGCGTCGGGCGGTGTGAGCGCGTTCGGT
CTGGGTGGCACCAACGTCCACATGATTTTGGAGGAGTTTCAGCGCGAACCGGCGCGAACAGC
GCGCGCACGCGCCACCTGACGGTGTGACGGCGCGGTGCGCGCAAGCCCTGGCGCAGCTGGCG
GCCAACCTCGCCGAACACCTGCGCGAACACCCGAGTTGGCGCTGGCCGATGTGGCCCATACG
CTGTGCGACGGCCGAAGCCACATCCATTGCGCGCATCCTGGTGGCGACCGATACGACGGCG
GCGATCGACGCCTTGATGAACGACCGCGATCCGCGAACGCGTTTCTTCGAAGCGACCGGGCGC
GGCGAGTCGGTGATCCTGTGTTTTGACGAAACGCCGCGGAGCCGCGAAGCGCCCCGCTACCTC
TGGGATCACGAGCCGCTTTATCGCGCGCGCGGCGACGTCGTGCTTGGCTGGTGAGGTGCGCGAC
CCGATCTGGAAGGCTGCTTTACTGCCCTGATCGCCGAGCAGGGCGCGGAGCCGCTTTTGC
CACCAATACGCGCTGGCCGGATGGCTGCTGGCCATGGGGTTGACCCCGTCGGCGTTGATCGGC
GTGGGCCAGGGCGAGTGGGTAGCAGCGGCGCTCGCGGAGGTGTTCCCGCCATCGGCCTGCTTG
CGCTGGATTAGGTTCCGGCGAACGGCTCCCGCAGCCGCGCGATCAACGGATTCCGTTTCTCTCC
AATTCTCTGGAACTGGATCGTTGGGCGTGAGTTGGCCGACCCGATTACCCAGAAAGCAG
AAGGGTAAGCGCTGCATGAAGCGCGCTCGGTCCCAACCTCGGTCAGCTGGTGCAGGATGGGGG
CGATGGAACCGGCTCGGTGAGCTCGTTCGCGCGCTGCTCTCCGCGGGAAGCGGAGCGGGGACG
GTGATCGGCCCCGAGGGCGAGGTTTCATCTCTGTCGACGAGCCGGGCGGGGTGCGCGCCCAG
TACCTGGGGGCGAGCTCGAGG

(2) peptide sequence

Seq ID No 46 (>pEP0cos6_ORF6.pep)

MACRFPGARNLAEYWANLIEGLETLSPFFSEELREAGCDPVQLAQHNYVRTKGLLPDADRFDA
DFFGYSPREAVMDPQIRVFHEVCWQALEHAGYNPHRHTGTIGLFAGAAPNVFWEFLSYRSDA
ANLGNFTLGLHNNKDYLSRIAENFNLTGPSYTLFTACSTSMVAIHQAVQALLNGECDLCMAG
SVSITLPLVAGYTYTPGMIVSPDGHCRTFDAGANGTVYGDGAGVVVLKRAEDALADGDHIFAL
IKGSALNNDGSRKTGYTAPSVQGGQVEVIRAAMNLAEEVEPEAISYVETHGTGTTVGDPLEFEAL
KEAFGGGCKAFPCGLGSVKPNIGHLDVTSGIASFIKLVLALEHRILPPTLHFQLPNPKMDVVDS
PFYIVAEREPPWRELLPRRAGVSFAFGLGNTNEMILEEFQREPAANSARTRHLTVLTARSPQA
LAQLAANLAEHLREHPALADVAHTLLHGRKPHPFARILVATDTTAAIDALMNDRDPTRFF
EATGRGESVILCFDETPEPRRSARYLWDHEPLYRAAATSCLAGEVADPDLEGCFALIAEQGA
AAAFCHQYALAGWLLAMGLTPSALIGVGQGEWVAALAEVFPPSACLRWIRFGERLPQPRDQR
IPFLSNFSGNWIVGRELADPDYPRKQKGRCKMRRRSQPRSAGAGWGRWNRLGQLVARCSSAG
SGGGTVIGPRARFISSTSRARVRAQYLGASSR

pEP0cos6_ORF7 sequences:

(1) nucleotide sequence

Seq ID No 47 (>pEP0cos6_ORF7.seq)

ATGGAACCGGCTCGGTCTAGCTCGTCGCGCTGCTCTTCCGCGGGAAGCGGAGGCGGGACGGT
GATCGGCCCGAGGGCGAGGTTTCATCTCGTCGTCGACGAGCCGGGCGCGGGTGCGCGCCAGTA
CCTGGGGCGAGCTCGAGGTAGCGGTCCCGCGGCCAGTAGGGCATCGCGGAATGACGTCCGC
CAGGTAGGCCTCCGGGTCGAGCCCGTGCAGCTTGCACTCGCCACGAGCGAGAAGAGGTTGGC
CGCGGCGGAGGCGTGGTCGTCGCTGCCGAAGAAGAGCCAGGACTTTCTCGCAACCGCAATGGA
TCGACGCGCTCGCTCGCTGGCGTTGTTCTCCAGGCGCAGCCGACCGTCGTCGAGGAAGCGCCG
CAACGGCTGCTCTTGTTGAGGGCGTAGCCGAGCGCGGTGGAGACCAGGCCGCGCTCGCGGGG
ACGAGCGTGCTCGGCCCTGGCCCAGGCAAAGAACGCGTCGACCAGAGGGCGGACGACGACATC
GCGACGCACCTTGCGCTGCGCGGCGCGCAGGTCCGCCAGCGCGCATCGGCGGCAAACAGGGC
GTTGATGCGCCGACGCCCTCGACACCGAGCTCGTGCTTGACAGACCGCCGCTCCAGAAAGTT
GGTACGGCAATGCGACCAGCATCCGACTTCGGTCGGGGCGGACCGCGCTTCTCGTCGGCAGC

5

10

15

20

25

30

35

40

45

50

55

AGCGCCTCTTGGTGGTGTGCCGCGGAAGAGGGCGTCATAGATGGCGTGAGCGTCAGCTTGAAT
ATACCGAGAGAAGCCGCGGAACATCTCGCAGACCGGGCGCTGGTATGCTTGGGCTGGTACTC
GAAGAAGACGTGATCCTTGTCCGCGAGGACGACGAAGAAGTGTCCCTTGCGGCACGGCCCCGGG
CTTCTTGTCTTGGCGCTCCTGGATGGGCCCAGGCTGGACGGAGACCCCGGTGGCGTCCGTGGA
5 CAGGCAGAAGCGGTCTCGAAGGCCTCTTTGCGCGCGGCCTCGACGATGGCGCCAGGGTCGC
ACCGACGTCTTCGGCGTAGCGGCACATCGTGCCGCGATCGAGCGACGCGCCCTGAAGCTCCAG
15 CTGCTGCTCCAGTCGATAGAACGGGACGCCGAGCAGGTAAGTGGTGGTGAGGATGTGCGCAAT
CATCGACGGCGCGAGGAACGACCCCGGAACAACTCCTTCGGAAGCGGCGTCGTGATGAAGAC
CGTGACAGTCTCGCCCTTCGGCGCGCGCGCGCGCTCGAGCGAGGGCGCGTCGAGCGCTGT
20 GGAGGAAGCGCTCGGCTCGCCGGCGCTGCCGTGTCCTCCGGCCTGACGCTCGCAGCGGGCGT
CGGCGTCGGGGCTTCTCTCGCGACGACCTGGAGCGGGGCGCTTCTCCTCGCCCCGAATGCT
CGCATCCGTGACGGACCGCTCGGCCTTGATACACGACGCGTGCGAGCACGATGCGGCGCATTCC
15 GCGCGCTCGTAGCCGAGTCGCGAGGTCTCCTCGACCCCGATGCGCGTCGCCGTGCGATCGAG
CTCGGGGCGAGGAGCTCGATGCGGACGACGGGCAGGTGCGACTCGGACAGGTGCGGACGGCC
CTTGCCGCGCGACCTTCGTTTCGGCCCCCTTGGGGTCGTGCTGCTGCGGCTCGTGCCTGTATT
GCGCTCGGCGGCGTCGAGTGCCCTTCGCGAGGCGCTGGACCTCGAGGAACATCGAGTCGAACGC
30 CAGCTGCTCCGCGCTCACCTCGGCGCGCTCCGCTTGGCCACGAACAGTCGACGTGCGAGAAG
CTGCAGCTGCTCGAGCGCACGGGTGTAGGCGCGCCGAAGCTGCGCGAGCGCATCGCGGCTCC
CACGAGCTCGCTCTTTGCCGCGGCGAGCTCCGCTTCGAGCTGCGCGATGCGCTGCTGCTCGGC
20 CGAGAGCGTCGGCTTGGCGGCGGCGTCGTGCACGACGCCGCTCTACGTAAGCCGCGCGTACTT
GTCGAGCGAATTTCGTGCGGCTCAGTGGACGCGGCGCGGTGCGCGCCTTCGCGGTTTGGACGTG
GGCGCGATCTCGATGCCGTGCGAGCAGCGTCTCGAGCGTGGCGTCGTCCACCTCGACGTGCGTG
GCGCCCTCGGTGCGGGGGTCCGGAAGTGCGAACGCTCCGCGATCAAGGCGTTTTGAAAACAGG
40 CAGATTCCACTGCCATCGAAGAAGAGAACTTGATCGTGGTCCGCCGCTTGCCGACGAACGCG
25 AACAGCGCTCCGCGAGCGAGCTCGTACCCACACGCTCACGGATGAGACCCGAAAGCCGCTCG
AAGCCG

(2) peptide sequence

Seq ID No 48 (>pEPOcos6_ORF7.pep)

MEPARSARRALLFRGKRRRDGDRPEGEVHLVVDEPGAGARPVPGGELEVAVPRPVGHRANDVG
QVGLRVEPVQLAARHEREEVGRGGGVVAAEEEPGLSRNRNGSQRS LAGVVLQAQPTVVEEAP
5 QRLLLVEGVAERGGDQAALAGTSVLGPGPGKERVDQRADDDIATHLALRGRQVRQRAIGGKEG
15 VDA PQPLDTELVLADRR LPEVGTAMR PASDFGRGRTALLVGSSASWWCAAEECVIDGVS VSLN
IPREAAEHLADRGAGMLGLVLEEDVILVREDDEEVSLAARPGLLV LALLDGPRLDGD PGGVRG
QAEGLLEGLFARGLDDGAQGR TDVFGVAAHRAA IERRALKLQ LLLQS IERDAEQVLAGEDVRN
HRRREERPPEQLLRKRRRDEDRAGLALRRRRRRRVERGRVERCGGSARLAGRCRVLRADARSGR
20 RRRGFSRDDLERGRFLLARTARIRDG PLGLVHDACEHDAHSAALVAESRGLLDPDARRRRRIE
10 LGAGELDADDGQVGLGQVATALAAGPSFRPLGVVVLPLVACIALGGVECLREALDLEEHRVER
QLLRAHLGALRLGHEQSTSQKLQ LLERTGVGAPKLRERIARSHELALCRGELRFELRDALLG
25 RERRLGGGVVHDAALRKPRVLVERIRAAQWTRRGARLRGLDVGAISMPSSSVSSVASSTSTCV
APSVGGSGSANAPRSRRFENRQIPLPSKKRILIVVRLPTNANSAPQRASYPTRSRMRPESRS
15 KP

pEPOcos6_ORF7.1 sequences:

(1) nucleotide sequence

Seq ID No 49 (>pEPOcos6_ORF7.1.seq)

ATGTTCTCTCGAGGTCCAGCGCCTCGCGAAGGCACTCGACGCCGCCGAGCGCAATACAGGCGAC
35 GAGCGGCAGCACGACGACCCCAAGGGGCCGAAACGAAGGTCCGGCGGCAAGGGCCGTCGCGAC
CTGTCCGAGTCCGACCTGCCCGTCGTCCGCATCGAGCTCTCCTGCCCGAGCTCGATGCGACG
40 GCGACGCGCATCGGGGTCGAGGAGACCTCGCGACTCGGCTACGAGCGCGGCGGAATGCGCCGC
25 ATCGTGCTCGCACGCGTCGTGTACAAGGCCGAGCGGTCCGTCACGGATGCGAGCAGTTCCGGGC
GAGGAGGAAGCGGCCCGCTCCAGGTCGTGCGGAGAGAAGCCCCGACGCCGACGCCCGCTGCG
AGCGTCAGCCCGGAGGACACGGCAGCGGCCGGCGAGCCGAGCGCTTCCTCCACAGCGCTCGAC
45 GCGCCCTCGCTCGACGCGCCGCCCGCGGCCGAGAGGCGGAGACCTGCACGGTCTTCATCACG
ACGCCGCTTCGGAAGGAGTTGTTCCGGCGGTTCGTTCTCGCGCGTCGATGATTGCGCACATC
30 CTCACCAGCAAGTACCTGCTCGGCGTCCCGTTCTATCGACTGGAGCAGCAGCTGGAGCTTCAG

5

10

15

20

25

30

GGCGCGTCGCTCGATCGCGGCACGATGTGCCGCTACGCCGAAGACGTCGGTGCGACCCCTGGGC
GCCATCGTCGAGGCCGCGCGCAAAGAGGCCTTCGAGACCGCCTTCTGCCTGTCCACGGACGCC
ACCGGGGTCTCCGTCCAGCCTGGCCCCATCCAGGAGCGCAAGGACAAGAAGCCCGGGCCGTGC
CGCAAGGGACACTTCTTCGTCTCCTCGCGGACAAGGATCACGTCTTCTTCGAGTACCAGCCC
5 AAGCATACCAGCGCCGCGGTCTGCGAGATGTTCCGCGGCTTCTCTCGGTATATTCAAGCTGAC
15 GCTCACGCCATCTATGACGCCCTCTTCGCGGCACACCACCAAGAGGCGCTGCTGCCGACGAG
AAGCGCGGTCCGCCCCGACCGAAGTCGGATGCTGGTGCATTGCCGTACCAACTTCTGGGAG
GCGGCGGTCTGCAAGCACGAGCTCGGTGTCGAGGGGCTGCGGCGCATCAACGCCCTCTTTGCC
GCCGATCGCGCGCTGGCGGACCTGCCGCCCGCGCAGCGCAAGGTGCGTCGCGATGTCGTCTC
20 CGCCCTCTGGTCGACGCGTTCTTTGCCCTGGGCCAGGGCCGAGCACGCTCGTCCCCGCGAGCGC
10 GGCCTGGTCTCCACCGCGCTCGGCTACGCCCTCAACCAAGAGCAGCCGTTGCGGCGCTTCCTC
GACGACGGTCGGCTGCGCCTGGAGAACAACGCCAGCGAGCGAGCGCTGCGATCCATTGCGGTT
GCGAGAAAGTCTTGGCTCTTCTTCGGCAGCGACGACCACGCCTCCGCCGCGGCCAACCTCTTC
25 TCGCTCGTGGCGAGCTGCAAGCTGCACGGGCTCGACCCGAGGCCTACCTGGCCGACGTCATT
15 CGCGCGATGCCCTACTGGCCGCGGGACCGCTACCTCGAGCTCGCCCCAGGTACTGGGCGCGC
ACCGCGCGCCGGCTCGTCGACGACGAGATGAACCTCGCCCTCGGGCCGATCACCGTCCCGCCT
CCGCTTCCCGCGGAAGAGCAGCGCGCGACGAGC

(2) peptide sequence

35

40

45

50

55

20 Seq ID No 50 (>PEPocos6_ORF7.1.pep)
MFLEVQRLAKALDAAERNTGDERQHDDPKGPKRRSGGKRRDLSES DLPVVRIELSCPELDAT
ATRIGVEETSRLGYERGGMRRIVLARVVYKAERSVTDASSSGEEEAAPLQVVAREAPTPTPAA
SVSPEDTAAAGEPSASSTALDAPSLDAPPPAPKGETCTVFITTPLPKELFRRSFLAPSMIAHI
40 LTSKYL LGVPFYRLEQQLELQGASLD RGTMCRYAEDVGATLGAI VEAAARKEAFETAFCLSTDA
25 TGVSVQPGPIQERKDKKPGPCRKHFFVVLADKDHVFFEYQPKHTSAAVCEMFRGFSRYIQAD
AHAIYDALFRGT PPRGAAADEKRG?PPTEVGCWSHCRTNFWEEAVCKHEL GVEGLRRINALFA
ADRALADLPPAQRKVR RDVVVRPLVD AFFAWARAEHARPRERGLVSTALGYALNQEQLRRFL
45 DDGRLRL ENNASERALRSIAVARKSWLFFGSDDHASAAANLFSLVASCKLHGLDPEAYLADVI
RAMPYWPRDRYLELAPRYWARTRARLVDDENMLALGPITVPPPLPAEEQRATS

5

116

pEPOcos6_ORF7.2 sequences:

10

(1) nucleotide sequence

Seq ID No 51 (>pEPOcos6_ORF7.2.seq)

15

5 ATGATTCCGGCGGGCGTGCAGGTGTTTCGTCGCGCTGGAGCCGGTGGACATGCGCTACGGCTTC
GAGCGGCTTTCCGGTCTCATCCGTGAGCGTGTGGGTACGAGGCTCGCTGCGGAGCGCTGTTT
GCGTTCGTCGGCAAGCGCGGACCACGATCAAGATTCTTCTTCGATGGCAGTGGAACTGTC
CTGTTTCAAACGCCCTTGATCGCGGAGCGTTCGCACTTCCCGACCCCCGACCGAGGGCGCC
ACGCACGTCGAGGTGGACGACGCCACGCTCGAGACGCTGCTCGACGGCATCGAGATCGCGCCC
10 ACGTCCAAACCGCGAAGGCGCGCACCGCGCCGCGTCCAC

20

(2) peptide sequence

Seq ID No 52 (>pEPOcos6_ORF7.2.pep)

25

MIPAGVQVFVALEPVDMDRYGFERLSGLIRERVGYEARCGALFAFVGKRRTTIKILFFDGSIG
15 LFSKRLDRGAFALPDPPTEGATHVEVDDATLETLLDGIETIAPTSKPRRRAPRRVH

30

pEPOcos6_ORF7.3 sequences:**(1) nucleotide sequence**

Seq ID No 53 (>pEPOcos6_ORF7.3.seq)

35

ATGACAAGGACGAAGGCGACCGAAGTGATGTGGTCCGAGCGCGTTCGGGCGTGGCGCGAGAGT
GGTGAAACGGCGGAGGAGTTCGCTCGGAGCCGCGATTTCGGCCTCGACGCTGCACGGCTGG
TCGAGCCGGCTGTCGCGGGCCGAGCCACCGCGCTTCTGCGCCTGGTGCCGAAGGCGCCCGCC
40 GTCACGAGCAGCGCTGCGGAGCTCGTCGTCGAGGTCGGCGGCGCGGGGTGCGCGTCGCCGCG
25 GGTTTCGACCCCGCGCTGCTGGCGGAGGTGGTCCGTGCCCTCGGCGGAGCGGGGCCA

40

(2) peptide sequence

Seq ID No 54 (>pEPOcos6_ORF7.3.pep)

45

MTRTKATEVMWSERVRAWRESGETABEFARSRGFAASTLHWSSRLSRAEPPRFLRLVPKAPA
30 VTSSAAELVVEVGARVRVAAGFDPALLAEVVVALGGAGR

50

55

5

10

pEPOcos6_ORF8 sequences:

(1) nucleotide sequence

5 Seq ID No 55 (>pEPOcos6_ORF8.seq)

15

20

25

30

ACTGGACAGCGCAGCCGGGGTGAGACGGCGCTTCGCGCAGCGCTTACGCAGAAGGCGCGCCGC
GCGCCATTGTCGGATGCGGTGCGCGACTTCGCCGCCGATCGGCTGTTGCTGGAAGTGGGACAA
CCACTGGACGTAACGGCTGAAGCGAGCCAACGGCTCCAGCTCCGCGGGGCGACCTGTTCCGGC
GCCTACCAAGCGTTGGCCCAGCTCTGGATCTGCGGCGCCCTGGCCGAACCGCCGCGACTGTAT
10 CCGGACGAACACCGCCGGCGCGTGCCTGCGGAGCTACCCCTTCGAGGGAAAGCGTTCTGG
ATCGAGGGCTCGCCGTTTCAAACCGCGCCCGCCGGCGCCTCACCCCAACCGCCGATTTCG
GGGGACATTCTCAAGGGCGACCCGGCGGACTGGTACTATCGGCCGCGTTTCAAAGCGGCGCCG
CTCTTGCCAGCCCGTTTCGAGAGCGAACCCGGCGATTGGCTGGTGTTCGAAGATGAGCTGGGG
25 CTGCGCGCCTGGCTGAGCGAGACCTTGCGCGACAAGGGCGCGCGGTTCGCGACAGTCGTTCTGA
GGCACCGAGTTCCGACGCCTGGCGTCACAGCGCTTCCAGCTTCGTCCCGATCGACGGGACGAT
15 TACCGGACCCTGCTGCACGAGTTGAAGGCGCAGGGCATCGCGCCGGTCCACCTGTGCCACCTA
TGGAGCCTGACCGCCGACCGGATGCCGAGCAGTTGCTCGACGTCAGCTTTCACAGCCTGGTC
30 CATTTGGCGGCCGCTTTGGGTTTCGGTTGGCTACTTCCACGCCATG

20 (2) peptide sequence

35

Seq ID No 56 (>pEPOcos6_ORF8.pep)

40

TGQRSRGETALRAALTQKARRAPLSDAVRDFAADRLLLELGQPLDVTAEASQRLQLARGDLFG
AYQALQLWICGALAEPPRLYPDEHRRRVPLPSYPFEGKRFWIEGSPFETAPAAGASPPADS
40 GDILKGD PADWYYRPRFEAAPLLPSPFSEPGDWLVFEDELGLGAWLSETLRDKGARVATVVR
25 GTEFRRLASQRFQLRPDRRDDYRTLLHELKAQGIAPVHLCHLWSVTAAPDAEQLLDVSFHSLV
HLAAALGSGVGYFHAM

45

50

55

5

118

pEPOcos6_ORF9 sequences:

10

(1) nucleotide sequence

Seq ID No 57 (>pEPOcos6_ORF9.seq)

15

20

25

30

35

40

45

50

55

5 ATGAAGTTGAACGTGGTCGCCAACCGGCTATTGACCCCGAGTCGCCCAGCGCACCGAGCCC
GCCAAGAGTCTGTTGCTCGCGGTGACCAAAGTCTGCGCGAAGAGGTGCCCAACGTTTGAACC
CGCGCCATCAGCGTGGACCTGGATCGCTCGTTGACGCGCGCGCGCCCGCTGGGCCCGCAGT
TTGTTGGTTGAATGCGGCGCGCCCGTCCAGGAAACGGTGGTGACCTACCATGGCGCAGCCCGA
TGGCTGCGCCGCTTCGATCGCGTTGCGGTGAATGGTCTCGGCCCGTTCCACCCCGATCAACCT
10 GCGCCGCTGCTGCGCGAGCGCGGCGTGTACCTGATCACCGCGGCGCTGGGCGGCGTGGCTGGC
CAGTTGGC3CGCTACCTGGCGCGGCGCTGCCGGGCGCGGTTGGTGCTACCGCGCGCCCGCCC
CTGCCCCGAGCGCGACCAAGTGGGATCGGGAGTCGGCCGTGCTGTCATGGGACGACAAGACGCGC
CAGCGCATCGAGCTGGTGCGGAGCTGGAGCGGCTGGGGGCCGAAGTATTGGTGGTGGCTGCC
GATGTCCCGACGAAGCGGCCATGGCGCAGGCGATCGAGGCCTCACTGGCGCGATTTCGACGCT
15 TTGGACGGCTTGATCCACGGCGCCGGGATCGTGCGGGTCGCGTCGGGCCGACGCCGATCGGG
AGTATGACGCGGGCCATGTGCGAGGAGCAGCTCCGCCCAAGATGTGGGCCTCGACGTCGTC
GACCGCCTCTGCGCGATCGCCGGTTGGAATTCGCGATTGCCATCTCGTCGCTCGCCCCGATT
CTCGCGCGCCTCGGCCACGTGCGCTACGCGCGCCCAACCTCTACATGGACGCGTTTCGCGACG
CGCGCCCGCCCGGCAACGCGCCTTGATCGCGCTGAACCTGGCCGAGTGGGAATACGAGGGC
20 CCGGCTACCTACGACGAGCGGGTGGGCCGTTTCGCTCAAGCAGCTCGAGCTCACCAACGAGGAG
GGTATCCGCGTCTTCCAGACGGTGTGGCCTTGCGCGCGCGCGCCCGCTACAGCAGATCATT
ATTTCCACCGGCGACCTCCAGGCCCGCTCGACAAATGATTACATCAAATCCCTGCATCGC
CGACCGGGGCCGTCCAGCTCAGTCGCGGACCGCGGCACCCCGAGGGCGGTTTCGGCTCGGAG
CGCGCCGCTTCGAGGCCGCTTCGCTGACGCCTGGTGCGACTTCTTCGGGGTTGAAGAGGTC
40 GACCCGAACAAAACTTCTTCGATCTGGGCGCCAGCTCGCTCGACTTCATCCACCTCGTCAGT
CGCTTCAGCAAGGCCATCGAACAGCATGTACCGCTCGAGGCCCTGCTCGAACACTCCACCCTG
CACGACCTCGCCGCCACCTCGCGGGCGACGCGAACACCGACGCCAGCGACGAAGCGCGCATT
CGCCAACGGCTGCAAGGCGCCAAGTCCGGCGACATCGCCATCATCGGCATGGCCGGCCGCTTC
45 CCGCTCGCGCCCGACCTGGACACCTATTGGCGCAACCTGGTTCGAGGCATCGACGCGGTTCAGC
TTCTTCAGCGCCGAGGAGTTGCGTCTGCTGGCGTCAACCGGCCGAGATCCACCACACCAAC
50

5

10

15

20

25

30

35

40

45

50

55

TACGTGCCGGCCAAGGGGCGCTGC3CCGACCAGGACTTGTTCGATGCGGCCTTCTTCAATAC
ACTGCCAGCGACGCCGAGCTGATGGACCCGCAAAATCGCGTGTACACGAGGTCGTGTGGCAC
GCGCTGGAAGACGCCTGTTTCGACTTCAACGGCGATCACGGCCAGGTCGGCCTGTTTCGCGGGC
GCCTCGCCGAACCTGTGGTGGCAGTTTCGTGGCCAGCTTTTCCGAGGCCGCAAGACGCAGGGC
5 ATGTTTACCACCACCCTGCTCAACGACAAGGACTCGATCGCGACCCAGATTTCATACAAGCTC
GGTCTAAAGGGCCCCGCGGTACCTTGTTCACCGGCTGTTCCACCTCGCTGGTAGCCGTTGAC
15 GCCGCTGCCGCTCGATCTGGTCCGCTCAATCGGACATGGCCGTGGCCGGCGCGGTCTCGCTG
ACTCTCCCCGATAAGGCCGGCTACATCTACGAAAAGGGCATGCTCTTCTCGGCCGACGGCCAT
TGCCGGGCTTTTCGACGCCAACGCCACCGGCATGGTCTTCGGCGACGGCGCCGGCGCGATCGTG
20 CTCAAGCCGTTGGACGCGGCCCTGCGCGACCGCGACCCGATCCATGCGGTGATCAAGGGCTGC
GCCACCAACAACGACGGCGACCCGAAAGCCGGCTACACGAGCGTCAGCGCCCAAGGCCAGGCC
GAGGTGATCCGCTCGGCCCAGATCCTGGCCGACGTGGCGCCCGAATCCATCAGCTACGTGGAA
25 GCCCACGGTACCGGCACCAAGTTGGCGGACTCGATCGAGATCAAGGCGTTGAAGCAAGCCTTC
GCCAGCGACAAGAACGGATTTTTCGGCATCGGGTTCGGTCAAGACCAACCTCGGTACCTGATG
15 GCGGCGGCGGGGATGGCCGGCCTGATCAAGACGTTCTGGCGATGAAGCACCGCCAATTGCCG
CCATCGCTGCACTGCGACGAAGTGAACCCCGACCTGGAGTTGGAGCGCAGTCCGTTCTACATC
30 AACACCCGCTGCGCGACTGGGTTGCACCGGGCGGGCCGCTGCGGGCCGGCGTGAGTTTCGTTT
GGGATCGGCGGAACCAACGCTCACGTATCCTGGAGGAGCCCGACGCGCGAGAGCGGCACG
CGCATGCGCCACTGGAAATTATTGATGCTGTGCGCGGCCAGCGAGGCGGCGCTCGACCGCCAG
20 GCGGATAACCTGGCCGACTACCTGGAGCGCCATCCCGAGGCCACCTCAGCGACGTGGCCTAT
35 TCCCTCCAGACCGGCCGGCGGCTTCTGGCCTGGCGGCGCACGGTCTATGCGAGTACCGCGAG
GACGCGGTGACCAGTCTGCGCGAGCGACAGGCCAAGCGGTCCAGACAAGTCGCGTCCGCTGG
GACCACAAGGACGTGGTCTTCATGTTTCCCGGTGAGGGCGCCAGTACCTCAACATGGGCCGCG
40 GACTTATACGTATGGAGCCGGTCTTCCGCGAGGTGATGGACCGCTGCTTCGAGTTGCTGGCC
25 CTTTGTGGTCCGAGCATCCGCGCCAGATCCTTTATCCGAGGGCGGGGTGTCGACCCTGCTC
CACCGGACTGATTACACCCAGCCGATCGTGTCTGCTTCGAGTACGCCCTCGCCCATTTGCTG
CTCTCCTGGGGATTGAAGCCGGCCGCGACCATCGGCTACAGCTTCGGCGAGTACGTTTCTGCC
45 TGCCCTCGCCGGCGTCTTCTCCCTGGAAGATGCGATCCGTCTGGTGACCGAGCGCGGTTCGGCTG
ATGGCGGCTTTGCCCCGGGGCGCCATGCTCAGCGTCCCGGTTCCCGAATGCGAGCTGCTGCGG
30 CTGCTGGACGGCTTCCACGCCCAATCGGCGGCCCATCTGGCGCTGGCCGTGACAAATGGCGCC

5

10

15

20

25

30

35

40

45

50

55

TCCTGCATTGTGGCCGGCGAGCAGGCCGCCATCTCGGCCTTCGAATCGATGCTTCGCAAGAAG
CGTCTGTTGACCATGCGGGTCGCGGTACGCCACGCCGCTCATTTCGAGGTCATGACCGGCGCG
ACCGACGCCCTGCGCAGCATCCTGCGGAAGATCCCCCTCTCCGCGCCGACAATTCCTTCATT
TCCTGCGTCACCGGCACCTGGATCACTGCACAGCAGGCTACGGATCGCGAGTATTGGGTGAAC
CACATGTGCGGGACGGTGCAGTTTCGCGGCGGGTCTGACCGAGCTGGGTCAAAACCGCGAGGCG
GTGTTCTGGAAGTAGGTCCGGGCCGCGACTTGACGTTGCTGGCCACCGCATCCTGGCCGAC
AGCGCGGCCGTGTTTCGAGCTGGTCAAGGCGCCCGACGGCGGCGACGACGATGGGTTCCTCCTG
CTGGATCGATTGGCCAAGCTCTGGAGGCTGGGGATTTTCGATTGACTGGGCCGCTTCTACGCG
GATGAGCGGCGGCGGAACTCTCGCTGCCGGGATATCCGTTTCGAGCGGCGCGCTTCTGGATC
GAGGGCAACCCGCTGGAGATCGCCGCCGGCAGGCCCAATGTCCAGGGCCGCTGGTCAAGGCG
TCGGACATCGGCGCTTGGTTCTACGTGCCGCAATGGCGGCGGTCTGGTGCTCGCCGAGCCGGT
ACAACGGCGGCGGCGCCGCCGTACGGCGGAGCAGGCACGCGTCGTGACCGAGCTACGGCGG
GGATGCGCGTCCGGCCGCTTGGGCAGCGGGGCTGCGGACTGAATGGCGGTGCCCCGTCCGAG
CGTCCGAAAGAAAGTGTAGCGCCAGCCGGTTCGACCAGCGCAGCGGCGCAGACCGGCGCGGAC
TGCCCCACACCGACTGGGGAGCCAGCGGTGTGCCAAAGGACGGGGCCGAGCCGCGGCCGACC
TGGCTTATTTTCGCCGACGCCGGCGGATTGGCCGAATCTTTCGCCAAGCGGGTTCAGGCCCGC
GGCGAGAAGCTTTACCTGGTGGCTTCCGGCTCGCGCTTCGAGCGCCTGGCCGAGACCCGCTTC
CGCCTCGATCCCGGGGCCAAGTCCGATCACCGCTGCTTTCAAGGCGCTCGACGAGGCCGAC
ATCCTGCCGACCCACCTCCTCGACTTCCGCTCGCTTGACTGCGGCGGGCCGACGCCGACCCC
ATGGACCAGGCCGGCTTCTTCGGGCTGTTGCACCTGGTCCAGGCGATGGCAGAGGCCGGCTAC
AGCCATCCCATTCGGCTGCTGATCGTCAGTTGCGGCGTCTACGATGTCACCGGTGCCGAACCG
CTGCAGCCGGCGCGGGCCACGATGATCGGACCGGCTCTGTGCATCCCGCAACAGTATCCGCAC
CTCGAAACGAGCCATGTGGATTTGGGCGTGGTCCATGCCGACGAGCTCCACGCCGCGCGCCAG
CTCGACAGCCTACTTGCCGAATGCCTAAGTGCAACGGCCGAGCGCCAATTGGCGCTGCGCGGC
CGACACCGCTGGCTGCTGGACTACGAGCCAGTCCGCTTGCCGCCGCTCGACCCGGGCCGCTG
CCCTGGCGCCAGCGCGGGTCTACTTGATCACCGCGGTTTGGGCGGGATCGGCCGCATCCTG
GCCGAACACCTGGCCCCGACGACCTCGGCTCGCCTGGTCTTAATCGGCCGCGAAACCTGCCC
GACCGGACGACTGGGACGCTGGCTGAACCGCCCGCAACCGGTTCGACGCCACCCACGAACGG
CTGCTGCACAAGATCCGCGGATTCGCGATCTGGAAGCGCTAGGCGCCGAAGTCTGTGCTC
GCCCGCGACGTCGCCAACGAAGCCGCCATGCGCGAGGCTACGATCGCGCCGAATCCCACTTC

5

10

15

20

25

30

35

40

45

50

55

GGCACAAATCCACGGGGTGATTACAGCGCCCGCCTGATGGACGCGCAAAGCTTCTCAC'GATC
GACGCCCTCGACCACGACCTCTGCGCCCGCCAGTTTGAAGCAAAAATCCGCGGCGTCTGCGTG
CTCGACCGCGTTCTGGCCGACCGCACGCTCGACTTCTGTCTGCTGATGTCTTCCATCTCCACC
GTGCTCGGCGGCCTGGGCTATTTTCGGTTACGCCGCGGCCAACGCCTTCCTCGACGCCTTCGCC
5 CAGGCGCGCAGCCGCGACGCCGCTTTCCCTTGGCTTAGCGTGGCCTGGAGCGATTGGAAGTAC
TGGACCGAGCGCAAGATGGACAACGAGGTCCGCGCGCTCATCGACAGCCTCTCGATGGAACCC
GCCGAGGGCTTGAAGCCGTCACCCGCGTCTTGGCTTGGGGCAAGGCGCCCCACATCGCCAAC
TCGCCCCGTGACCTCGGTCCGCGCCGGGATCAATGGGTCAAATGGCCAGCCTGAAATCGGCG
CACTCCAGCGAGCCCGAGCCGGCTAGGCATGGACGTCCGCGCTCTCCAGCGAATGGGTCCGCG
10 CCGCGCAACGTGGTGAAGAGAAGCTGGTCGCCATTTTCGAGCAGGTGTTCCGGCACTGCGGCA
CTGGGCATCGAGGACAAC'TCTTTGAGTTGCGCGCGACTCGCTCAAGGCGGTTCATGACCGCG
GCCCGTATTCAAAGGAGCTGAACGTGGAAGTGCCGCTGCCGACCTTCTTCCAGATGCCACG
GTGCTGGCCTGGCCAGTTTCGTGACGCAAGCCAAGCGCAGCGCCGGGAGACGATTCCGCGC
25 ACCGCGCCGCGCCACATTACCCGCTCTCGGCTGCCAGGGCCGCCATTACCTGCAC'TACCGC
ATGGACCCGCGTTGTACCGCATACAACGATCCCTTCGCCAACCTGATCGAGGGTCCGCTGGAC
GTGGATCGCGTGGAGCGCATCCTGCACACCCTCATCTACGCCACGACTGCTTCCGCACCTCG
30 TTCCACTTCCGCGAGGGCGAGCCGGTCCAGGTGATTACGATCGGGTGGACTTCAACCTGGCG
CGGATTACCTGCGCGCCCGAGGATTTGCCCCGAACGGATGCGCGATTTCATCCGCTCCTTCGAT
CTGGAGCGACCGCCCGCCATGCGCGCCGGCCTCTTCGTACGGGGCCCGAGCGCCACGTGCTG
20 CTAATCGATTTTACCACATTATCACCGATGGCGTGTGTTTCGAGAACTTCGTGCGCGAGTTTC
35 GCGGCGCTCTACCGCGCGAGATCCTGCCCGAGCTGGAACTCGAGTACAAGGATTTTCGCGGTG
TGGCAGCATGAGAACCGGGGCCCGCCGCAACAGCGACCGAGCCCGCTACTGGACCGAGCAG
TTGGCCAATGCGCCCGGGCCGATCGAGCTAACCCGATTTCCCCCGTCCAGTCGACGCGAGC
40 TTCCGCGGCGACCGCGTGCAGGACCGTGTGTTGATGCGGAGCTCGTTGCTCGACTCAAAGAGCAC
25 GCGGCGCGCCTCGGCATCACCTCTATAGCCTGCTGCTGGGCGGATTCTCGTTATTGCAGCAC
AAGCTCTCCGACTCGCACGACATCGTCATCGGTTCCGCCGTCGCGGGCCGCACCCGAGCGAA
CTCCAGGATCTGCTGGGCGCGTTCGTCAACACCCTGCCGATGCGCCACCGCATCGACCCGACC
45 CATACCGCACGGGTCTTCTTGGAGCAGGTCCACCAGACAACCTTGGCGGCCCTCAGTACCAG
GAGCACCTTTTTCGAGAAATGGTGGCGACGCTCGGTTTCGCCGCGGATCCGGCTCGCAACCCG
30 ATCTTCGACACGATGTTCTTGCTGCAGAACATGCCATGGGTGCAACCACCATTTCCCGGTCTG

5

10

CGGCTCTCGCCTCACGACACTTTTCACCGCAAGGCATTGTGCGACCTGATGCTACAGGCGACC
GAGTATGACTGCCACCTGGAGCTGGTGCTCGAGTTCGCCACCGACCTGTTCCGGCTGGAACCC
GCGCAAGTCTTGCTCGACCGCTACCGCCAAGTCTTGAGTGGCTGTTGGCSTACCCCATGAA
TCGATAGACGATTTGACGCTCGCCGGCCACTTTCGCGAAGTCGAAGTGACGATGTGGACGAG
5 GGCGACTTTGATTTCTCAGATTTCAACCCCGCAACGTGAGAAACCTATGGCGCGCC

15

(2) peptide sequence

Seq ID No 58 (>pEPOcos6_ORF9.pep)

20

10

25

15

30

35

40

25

45

30

50

MKLNVANRLFDPESPERTEPAKSLLLAVTKVLPQEVNVRTRAI SVDLDRSFDAAAPAWAAS
LLVECGAPVEETVVTYHGAARWLRRFDRVAVNGLGPFHPDQAPAPLLRERGVIYITGGLGGVAG
QLARYLARACRARLVLTARRPLPERDQWDRESAVLSWDDKTRQRIELVRELERLGAEVLVVA
DVADEAAMAQAIEASLARFDALDGLIHGAGIVRVASGRTPIGSMTRAMCEEQLRPKMLGLDVV
DRLLRDRRLDFRIAISSSLAPILGGLGHVAYAAANLYMDAFATRAAAGNAPWIALNLAEWEYEG
PATYDERVGRSLKQLELTNEEGIRVFQTVLALAARGPLQQIIISTGDLQARLDKWIHIKSLHR
RPGPVQLSRRTAAPQGGFGSERAFAAFADAWCDFFGVEEVDPNKNFFDLGASSLDFIHLVS
RFSKAIEQHVPLEALLEHSTLHDLAAHLAGDANTDASDEAIRQRLQGA KSGDIAIIGMAGRF
PLAPDLDTYWRNLVGGIDAVSFFSAEELRAAGVTAAEIHHTNYVPAGRCADQDLDFDAFFEY
TASDAELMDPQNRVLHEVVWHALEDACDFDNGDHQVGLFAGASPNLWWQFVASFSEA AKTQC
MFTTTLLNDKDSIATQISYKLGLKGPVATLFTGCSTSLVAVDAACRSIWSGQSDMAVAGAVSL
20 TLPDKAGYIYEKGLFSADGHCFRANATGMVFGDGAGAVLKPLDAALRDGDPHIAVIKGC
ATNNDGDRKAGYTSVSAQCQAEVIRSAQILADVAPESISYVEAHGTGTLGDSIEIKALKQAF
ASDKNGFCGIGSVKTNLGHLMAGMAGLIKTVLAMKHRQLPPSLHCDEVNPDLELERSFFYI
NTRLRDWVAPGGPLRAGVSSFGIGGTNAHVILEEPPTRESGTRMRHWKLLMLSAASEAALDRQ
ADNLADYLERHPEAHLSDVAYSLQTGRRVLAWRRTVLCEYREDAVTS LRERQAKRVQTSRVRW
25 DHKDVMFMFPQGAQYLNMGRLDYVMEPVFREVMDCFELLAPLWSEHPRQILYPEGGVSTLL
HRTDYTQPIVFCFEYALAHLLLSWGLKPAATIGYSFGEYVSACLAGVFSLEDAIRLVTERGRL
MAALPAGAMLSVPVPECELLRLLDGFHAQSAHLALAVDNGASCIVAGEQAAISAFESMLRKK
45 RLLTMRVAVSHAHSQVMTGATDALRSILRKIPLSAPTIPFISCVTGTWITAQQATDREYWN
HMC GTVRFAAGLTELQONREAVFLEVGPGRDLTLLAHRILADSAAVFELVKAPDGGDDDGFL
30 LDRLAKLWRLGISIDWAGFYADERRRXLSPGYPFERRRFWIEGNPLEIAAGRPNVQGPLVKA

55

5

10

15

20

25

30

35

40

45

50

55

SDIGAWFYVPQWRRSVLAEPGTTAAGAAVTAEQARVVTEL RAGCASAGLGSGACGLNGGAPSE
RPKESVAPAGSTSAAAQGTGADCTPTGEPAAVPKDGAEPRPTWLI FADAGGLAESFAKRVQAR
CEKLYLVASGSRFERLAETRFRLDPGAKSDHRLLFKALDEADILPTHLLDFRSLDCGGPDADP
MDQAGFFGLLHLVQAMAEAGYSHPIRLLI VSCGVYDVTGAEPLQPARATMIGPALCIPQQYPH
5 LETSHVDLGVVHADELHAA RQLDSLLAECLSAERQLALRGRHRWLLDYEPVRLPPLDPGRL
PWRQRGVYLITGGLGGIGRILA EHLARTTSARLVLIGRETLPDRDDWD AWLNRPQPV DATHER
LLHKIRAIRDLEALGAEVLVLAADV ANEAMREAYDRAESHFGTIHGVIHGAGLMDAQSFSLI
DALDHDLCARQFEAKIRGVCVLD RVLADRTLDFCLLMSSISTVLGGLGYFGYAAANAF LDADA
QARSRDAAFPWLSVAWS DWKYWTERKMDNEVGAVIDSLSEMEPAEGFEAVTRVLAWGKAPHIAN
10 SPGDLGRRRDQWVKLASLKS AHSSEPEPARHGRPALSS EWVAPRNVVEEKLVAIFEQVFGTAA
LGIEDNFFELRGDSLKAVMTAARIQKELNVEVPLPTFFQMPTVAGLAQFVTQAKRSGRETIRR
TAPRPHYPLSAAQGRHYLHYRMDPRCTAYNDPFANLIEGPLDVDRVERILHTLILRHDCFRTS
FHFREGEPVQVIHDRVDFNLARITCAPEDLPERMRDFIRSF DLERPPAMRAGLFVTGPERHVL
LIDFHIIITDGVSFENFVG EFAALYRGEILPELELEYKDFAVWQHENRGRRANS DQARYWTEQ
15 LANAPGPIELTTDFPRPSRRSFRGDRVRTVLD AELVARLKEHAARLGITLYSLLGGFSLLQH
KLSDSHDIVIGSPVAGRTRSELQDLLGAFVNTLPMRHRIDPHTHTARVFLEQVHQTTLAALS YQ
EHPFDEM VATLGFAADPARNP I FDTMFLQNAMGATTI PGLRLSPHDTFHRKALCDLMLQAT
EYDCHLELVLEFATDLFRLETAQVLLDRYRQVLEWLLAYPHESIDDLTLAGHFREVEVTMSDE
GDFDFSDFEPRNVRLWRA

20

pEP0cos6_ORF10 sequences:

(1) nucleotide sequence

Seq ID No 59 (>pEP0cos6_ORF10.seq)

25 ATGGCGCGCCTGAGCCGCACAGATCTCCA ACTCGCCATTCACCAGCGCACCGTGGAGCGCGAA
TATTGGCGCGCTCTGTTCGAGCGCCATCCGCAACGGTCCAGTTTGCCGGGGGTGCTCACC GCC
CCGATCGGCGACGAGTCGACCCGCGAGACCTTGTCATTTCGTCCTCGACGAAGATCCCCTTCGG
45 CTGAGTAATCGTTCCCGCAACGCCTGCTCAGGTGTTGGCGGCTGGCCTCGCGGCTTTCTC
CACCCTGCGACGGCGCTGAGCGCTTCA CCCTGGGGTTGGCCCTACCGCGCCAAGCCGATGAC
30 CATCACCCGATCCTCAACAGCTTGATCGCGCTGGGGGTGCGGCTCGACTCGAGTACGACCTTC

5

10

15

20

25

30

35

40

45

50

55

CGCGATCTGCTCTATGCGCTTCGATCCGAATACCACGAGGCGATGCGCCACGCCAACTTTCCG
CTGGCGACCTGGTGGCGCGGCCTACCCGGCGGAACGGCGCCGTTTCGACGTGCCCCCTCAGCCTG
GACCCCTTCACAGACGGCGATTTCGCTGGAAGACCACGCGATCGGCGCGTTGTTCCGGTTTCGA
TTGGAGGGTGAGCGCCTCACCTGCCGATTGCGATTTCGACCTGCGCGCTATGACCGTCCC GCG
5 ATCGAAAACCTCGCCGATCGTTTCGCCCCGTTCTCAGCGCCTGTGCCGGGACGCCTCCACC
15 GTCATCCAGGCGCTGGACCTTTCGCTGCCAAGCGATGAATCGGTGTGGCGCGTCACTGAAGGC
GTGCGGCGCGGCTATTTCGAAGACCTGACGCTAGACCGCGCGTTCCGCCGCCAGGCCGCGCAA
ACGCCCCGATCAGCCGGCGATCACGTTGAACGGGGACGTCCAGAGCTACGCCGAGGTTCGACCGC
CGCAGCGACGCGCTGGCCCCGCCACCTCCGTCGCCACGGCGTCGGTCCGGAAACGATTGTGGCC
20 GTCAACCCCCGGCGCGGCCTAATCAGCTGACGGCCCTGCTCGCGGTCCATAAGGCCGGCGGC
GCCTACCTGCCGATCGATGCCGAGGAGCCGGCTGCCCGCCAGCAATTCAAGGTGCGCGACAGC
GGGCGCGGTTGGCACTGGAGCCGTGCGCGGACCAGGCGCTGACCGTCACCGACCTGCCGCGG
CTCTTCTGACGATGCCTCGCTCTTCGCTGACGGCGGGCTCGATGTGCCGCGCGGCGCGGAC
25 TCGCTCAATCCGGCCTATGTGATGTACAGTCCGGCTCGACCGGACAGCCCAAGGGTGTGGTG
15 GTTCCCCACCGCGCGTGGTCAATCGTTTGAATTGGGGGAGTCCCGTTTCCCGCTGGACGAA
CGCGACCGAATCCTCCAAAAGACGCCGTGCTGTTTCGACGTGTCCGTTCTACGAGCTGTTCTGG
30 GCGCATGGAGCGGGGCCACCTGGACATCCTCGAGCCCGCGCGGAGCGCGACCCCGACGCA
GTGGCCAGGGCCCTGGCCGAGCGCGCCATTACCGTATGCCATTTCTGTCCTTCGATGCTGCTC
GTCTACTTGGAAGTCATGCGGCGGCACCATGCGCCGCCCGTGGCCGACCGGCTCCGTTACGTC
20 TTCGTGAGTGGCGAGGCCCTCGAACCAGGACACCTCGCCGGGCTCCAGCAGATTGGTCGGCGC
35 CTCGGCCGCACGATTCCCCTCGTTAATCTGTATGGACCAACCGAGGCCTCGATCGAAGTCTCC
TGCTTCGCCTGTCCCGCCGACCATGTGCCGCGCCGGATCCCCATCGGGCAGCCGATCGACAAC
GTCGCACTGCACGTTCTCGACCGGCGCGGCCGTGCCAGCCGCCCTATCTTCTGCGGAGCTG
40 TTCCTGGCCGGCGACTGCCTGGCGCGCGGCTACCTCAACCGTCCCGACCTGACCGCGCTCCAC
25 TTCGTGCCCAATCCCTTCGGCAACGGCGAGCGCATGTACCACAGCGCGGACTTGGCGCTCGTG
CGCGCGACGGCCAAGTGGCGTTTCTCGGCCGCCGTGACCACCAATCAAAATCCGTGGTCAA
CGGGTCGAAC TGCGCGAAATCGAGAGTCATTTGCGCGGGCTCGAAGGCATCGCCGCCCGGTC
45 GTCCAGGCCGAGTCGCAGCACCATGAAACCCTGCTGCACGCCTACGTGCTCACCACGACGCG
GGCCTCAATGCGGCCCGGCTGCGCGCCGCCCTCGCTCAACATCTGCCCGAGTACATGATTCCC
30 CAGCGCTTCTCGCGGCTGGCCGAGTTGCCGCTGCTGGCGGCAGGCAAGATCGACCGCGCCGCC

5

125

10

5

15

CTCGCGCAACGTGCAACGCCGCTCGCCAGCGGCGGCCCCCTTCGTGGAACCCAGCGGGCCCCACC
CAGCAGCGTATCGCAGAACTGTGGCGCCAGGTCTTAGCGGTGCGCGAAGTCGGCGCCGAGGAT
CCCTTCTTCAGCATCGGCGGCAACTCGCTCAATGTGCTCAAGCTCAGCGCCGCGCTGAGCGAC
GCCTTCGCGCGTGACATTCCTCATGCCGGCCCTGTTCCAATACGACACCATCGCCGCCAGGCC
TCCTGGCTCGACGGGCAGGTTGACGAACGGGCCCAATCCGCCGCGCTCGACCGGCAGGCCCGCC
GAGGCGGCGCTGACCCTTCAAGAGACCGTGCCATTTTTGAGGGATTGATGACGAACCA

(2) peptide sequence

Seq ID No 60 (>pEP0cos6_ORF10.pep)

20

10

25

15

30

35

20

40

25

45

50

55

MARLSRTDLQLAIHQRTVEREYWRALFERHPQRSSLPGVLTAPIGDESTRETLSFVLDEDPLR
LSNRSPQRLLTVLAAGLAAFLHRCDDGAERFTLGLALPRQADDHHPILNSLIAGVAVDSSTTF
RDLLYALRSEYHEAMRHANFPLATWWRGLPGGTAPFDVALSLDPFTDGDSDLHAIGALFRFA
LEGERLTCRLRFDPARYDRPAIENLADRFAFLTRLCDASTVIQALDLSLPSDESVMRVTEG
VRRGYSQDLTLDRAFRRQAAQTPDQPAITLNGDVQSYAEVDRRSDALARHLRRHGVGPETIVA
VNARRGPNQLTALLAVHKAGGAYLPIDAEEPAARQQFKVRDSGARLALEPSPDQALTVDLPR
LFLDDASLFADGGLDVPRGADSLNPAYVMYTSGSTGQPKGVVPHRGVNVNRLNWGQSRFPLDE
RDRI LQKTPLLFDVSVYELFWGAWSGATLDILEPGAERDPDAVARALAERAITVCHFVPSMLL
VYLEVMRRHHAPPVPDRLRYVFSGEALEPDHLAQLQIGRRLGRTIPLVNLGYPTASIEVS
CFACPADHVPRRIPIGQPIDNVALHVLDRRGRROPYPYLPGEFLAGDCLARGYLNRPDLTALH
FVPNPFNGNRMYSGLDALVRGDGQVAFGLRRDHQIKIRGQRVELGEIESHLRGLEGIAAAV
VQAESQHHETLLHAYVVTNDAGLNAAARLRAALAQLPEYMIQRFSLAELPLLAAGKIDRAA
LAQRATPLASGAPFVEPSGPTQQRIAELWRQVLAVAEVGAEDPFFSIGGNSLNVKLKLSAALSD
AFARDIPMPALFQYDTIAAQASWLDGQVDERAQSAALDRQAEEAALTQETVAIFEGFDDEP

5

126

pEPOcos6_ORF11 sequences:

10

(1) nucleotide sequence

Seq ID No 61 (>pEPOcos6_ORF11.seq)

15

5 ATGACGAACCATGACCATCACGAGGAGAGCAGCGGCCTGGAGATCGCCGTCATCAGCATGGCC
TGCCGATTCCCGGGTGCTATGGCCTGCCGATTCCCGGGTGCTGCCGATTGCGACGCATTCTGG
GAAAACCTGATCAACGGGACCTCCTCGATCACCCATTTCAGCCACGACGAGCTGATCGCGGCC
GGCGTTGACGCGCGCGACCTGACGCGCAGTACGTGCGCGCGGCCGCCAGATCGATGACGCC
GAACGGTTCGACGCGGCCTTCTTTGGGTACTCCAGCGTGAGGCCGAGCTGATGGACCCCCAG
20 TTCCGCCTGCTCCATGAATGCGCCTGGTCCTGTCTGGAACAGGCCGGCATCGATCCGCGCGTC
GAAGCCGCGCCGATCGGGCTGTATGCCGGCGCAGCCGACAACACCTACTGGAACGCGCTCTCG
TCGCTCGACCGGGGCTCGGCCGAATCGGAGCAATTGCGCGCCGAACAACCTTGCAACCGCGAT
25 TTTCTGTGACGCTGGTCGCGCCGCGCTCAACCTGAAAGCCCCGCGGTGGTGGTTCAAAGC
GCCTGTTGACCTCGCTGTGGCGGTCCACTCGGCCTGTCTGCGCTCCTGACCGGCGAATGC
15 CGAGTGGCCTTGGCCGGTGGGGTGGCGCTGCGCTTCCCACGCCCAGCGGTTATCGCTACGAA
CCTGGCATGATCTTCTCGCCCGACGGGGTGTGCCGCGCGTTGACGCGGGCGCTAACGGGACG
30 GTGCCCCGCCAAGGCGCGGGGCTGGTAGCGTTGAAGACGCTGAAACGTGCCCTCCAGGACGGC
GACACGATCCACGCCGTGATTGCGCGGACCGCGGCAACAACGATGGTGCCCGCAAGACCGGG
TTCACCGCGCCCAGCGCCACGGCCAAGCCGAAGTCATTGCGACGGCGCTGCGCCTGGCCCCG
20 GTGCCGCGCAATCGATCGACTACGTGAGGCCCCACGGAACCGGCACGCCGCTAGGCGACCCG
35 ATCGAGGTAGCCGGCTTGGTGGAGGCCCTTCGCCAGCGAGAAGCGCGGCTATTGCCGGCTGGGC
TCGGTCAAATCCAACCTTGGTCATCTGGACACTGCTGCCGGCATCGCCGGCCTGATCAAGACC
GTGCTGGCGCTCGAGCACGCGCACATCCCCAAGTCCTGCCACGTGCCACGCCCAACCCCGCG
40 GCGCGCTACACAAGACGCCTTTCCGCATTGCCGCGACGGGATGGCCTGGCCGCGGCGTATG
25 GCGACGCCGCGGGCGGGCGGTTGAGTTCGTTCCGCATCGGCGGCACCAACGTCCACGCGATT
TTGGAGGAGCGCCGCCCGCGCGCCGAGCTGGCGGACGGGCGCAGTCAGGTGTTCTCTTC
TCCGCCAAGGACGAGGCGGCGCTGGACCGTGCCCTTGCCAACTATGGTGCGGCCTTGAGAAG
45 CGCGGCGACCTCGCGGCGGGCGCGGTGGCTGGACGCTCCAAAACGGCCGGGCGCATTCGAA
TGGCGAGCCAGCGCGGTGGCATCCGACCTCGACGAATTGGCGGGCGCATTCGCGGGCGAGCGG
30 CCCGGCGCGTCAAGAAAAACCGAATGGCGCGCAGGATAAGCCGGTGGCGTTCTTATGTTG

50

55

5

127

GGGCAGGGGAGCCAGTACCGTGGCATGGGCCACGACCTGTACCGCGAAGAGCCGCGTTTCCGG
CACCACCTCGACGCCTGCCTCGCCATCCTCGCCGAACACAAGCCCGAGATCGACTGGCTGGCG
TTGCTGGGCTACCGCGACGAGGACGAGCCAACCGACCAGATCGGGACGTCTCGCAGGGCCCG
AGCCGGTCAGCCGCATCGAACCCAGCGGAGCTCCTCGACAGCACCGAATTGCCCCAACCTTTG
CTTTTCTCCATGTCCTACGCGCTCGGTGCGCTGTGGCTCGACTGGGGCGTGCGACCCACGGCG
ATGATCGGGCACAGCCTGGGCGAGTACAGTGTGTCATGTATTGCAGATTCTATGCACTCGAT
CAGGTGCTGCCCTTCATTCTGACCCGCGGTGAGTCATGGCGCAATTGCGGCGCGGCTCGATG
TTGGCCGTGACGGTGACAGCGTTCTGATGCGCGAGCTGATCGCCGATGCGCTCGATTGGCG
GCGATCAACGGCGCTGACCAATTTGTCTGGAGCGGGCCGAGCGAGGCTGTCCAAGCCGCGGGG
GTCCGACTGCGCGGCGCCGCGCTGCGTGCCACCGAGCTGAACACCTCACACGCGTTCCATTCA
GCCATGATGGATCCCATTTCTGGAGSAGCTAACGGTTGCCGGTTCGCGACTTCAGGTGCGTGT
GGGACGATTCCGGTCGTTTCATGCGTTACCGGAACCTGGTTGACGGCGAAGCAGCTGGCCGAT
CCGCGCTACCACGCGCTCACGCGCGGAACCGGTGCGGTTGCGGCGGGCCTAGCGACGCTG
ACAGGGGAGGAGCCCGCTGATGCTCGAAGTGGGGCCGGGCTCGACCTGGCGGCTTTGGCC
CGCGAGCATTGAAATGCCCCGCTCCCGGTGCTCACCAGCCTGCGCCACGCTCGCCAGGCGACG
CCCGATCGCCAATACCTGCTCGAAACGCTCGGCTGCCTTTGGCGACACGGGTTTCCGTCGAT
TGGGGGGCCCATGCCGACGTTTCGCGACGCTTGTTTCGCTGCCCGGCTATCCCTTTTCCGGC
GCGGTGCGCCGCTTAGCCGGCGACCCCTCCGCTGCTGGCCGAGCCCGCGCGCTCGCCGCC
CCGTGCGGAACGCGCAACTCAGCGCCGACGCGCGGACCTCCCGAACACTCCGAGCCGACA
TCCGGCGCCGTGTCGGCGATCAAAGCGCCAATCGCCGCCGCGATCCCGGCCTCTATCGCCTC
TCCTGGCGCCAGGCCGGAACGGCGCCGCTCGGTCCGCCGATCTCGGTCCGCCCCGCGACTGG
ATCGTCTTCGCTCTGATTCTCACCTGCTCCAGGCGCTCAGGGCCAATCTCGGGACGCGCGCT
CAGCGGGTGACGCTGGTGACGCCGGGCCAGGAGTACGCAGCCGAGCCGTCCGGGTTTCGGCTG
CGGCCGGACCAGATCGACGATTACCGCGCCCTGTGGGCGGACTTGGCGCAAACCGGTATTGTG
CCACGATACATCGCGTTCCTCGCCCCGTTTCATGTACCGGGCGCGCATGGCGGGCGATGCCTCG
ACCCTGGACGAAGTGCGCGAGGGCGGCTTCCTGCCCCGTGACCCGCTTGATCCAGACTCGCCCCG
CCAGGGCGACCGAGCGGACTTCTAAGCCTCACGATCGTCACCCCGCGCCCTGGCGCTGGGC
GACGAAGCGACGCGCCCGGAATGGGCAATCCTGCACGGGATGGTCGCCGGCTTAAGCCGCGAT
TATCCCGAATGGCGCTTCGTCTCGATCGACGGCGCGACCCATCCCCGCATCGGTGCGAAGGT
CTGCCCCGCTTGATCGCGCTTCATGCGGTGACGAGGCTGGCCCCGACCCGCTTGGCGCTGCGC

50

55

5

10

15

20

25

30

GGCCTTCACGCTTGGGTTCACAGTGCAGCACGTTACGCCGGCCACCATCCCTGGGGCGGGT
ATGTGGCGCGAGGGTGGTGTGTACATGATAACGGGCGGATTCTGGCGGGATCGGTCTGGCGCTG
GCCCCGCGCCTGGCTCGAGAAGCTCGCGCCAAGCTGATCCTGGTCGGCCGAAACCTGCCCACC
CGCCGATCGATCTCGAGGCTTGGGACGCGCGCCGTTGATTCTCACCGCCGACGTGCGCGAC
5 GAAGAGGCCATGCGCCCGTCTTCGATGCCGCGCACGCCCGTTTCGGCGCCATCGACGGCATT
CTTCACGCGGCGGTCTCCCGGTGGCAGCCTGTTCCGCCAACCAATCGGACGCGGCCTTCGAA
15 GACGTGCTGCACGCCAAGGTTTCGCGGTACCCTCGTGCTGCAAGGCCTGAGGGCAATCGATGCG
CCGTGTTGCTGATGTCCTCGCTGGACGCTGGCTTCCCGGTCCCGGTGACACCGCTATGCC
GCCGCCAACGCCTTCCTCGACGCCTTCGCCAGTCTGCGCCGGCGAGAGGGAGAGCCGGTGATC
20 AGCGTTGGCTGGGACAGTTGGTGCAGGTGGGCATGGCTGCTCGGGTCGCTGCCCGATCGGCC
GACGAACGCGGCCGCTTGGCGCGCGAGGGGATCAGCCCTCGCCAGGGTTGGCAGGCTTTGAGC
CGGGCGCTCGCCCTCGACCCCCCCCCACCTGATGATCTCGCGCACCGACCTGACCTCGCGCTGG
CACAGTCGATCCAGCCCTACGCCGTCGCTCGAGCGAACCCGAGGTGGCGCTGCCGCGCTGG
25 ACCGCATCCGCCTGCCAAGCCGTATCGAGCGTGTGTTGGTGCAGCACTTCGCCACCGCCGCC
15 GTGCCTCCCGATGGCAACTTTTTTCGAGCTCGGCGCCAGTTTCTTCGACATCGTCCAGCTCAGC
GCTCGACTTCAACAACAGTTTCGGCCGAGATGTCAGCCACACCGTGCTCTACAGTCATCCACC
30 GTCGCCTTGCTGGCCGGCTACTTCGCCAATGACCCGACGCGCTCCGGTGCTGCTGCCGACGAA
CGCGACGAAGCGGTGCGTTCGCGGCCGCGACCTCTTGAAGAGCCCGCGGCGAGGAGTA

35

40

45

20 (2) peptide sequence

Seq ID No 62 (>pEPOcos6_ORF11.pep)

MTNHDHHEESSGLEIAVISMACRFPGAADCDAFWENLINGSSTHFSDDDELIAAGVDARDLT
PQYVRAAGQIDDAERFDAAFFGYSQREAEELMDPQFRLLHECAWSCLEQAGIDPRVEAAPIGLY
40 AGAADNTYWNALSSSLDRGSAESEQFAAEQLCNRDFLCTLVAAALNLKGPVVVQSACSTSLLA
25 VHSACRALLTGEGRVALAGGVALRFPRPSGYRYEFGMIFSPDGVCRPFDAANGTVPGEAGL
VALKTLKRALQDGTIHAVIRATAANNDGARKTGFTAPSAHQAEVIRTALRLARVPAESIDY
VEAHGTGTPLGDPIEVAGLVEAFASEKRGYCRGVSVKSNLGHLDTAAGIAGLIKTVLALAH
45 IPKSVCHVATPNPAARLHKTPFRIAADGMAWPRRMAPRRAVSSFGIGGTNVHAILEEAPRA
PELADGRSQVFVFSKDEAALDRALANYGAALKRGDLAAGAVAWTLQNGRAAFEWRAVAS
30 DLDELALALRGERPGAVKKNRMAREDKPVAFLCSGQGSQYRGMGHDLYREEPRFRHHLDACLA

50

55

5

129

10

15

20

25

ILAEHKPEIDWLALLGYRDEDEPTDQIGTSSQGPPSRSAASNPAELLDSTFAQPLLFMSYAL
GRLWLDWGVPTAMIGHSLGEYSAACIADFYALDQVLPFILTRGRVMAQLRRGSM LAVSGDSV
LMRELIADALDLAAINGADQFVWSPSEAVQAAGVRLRGAGLRATELNTSHAFHSAMMDPILE
ELTVAGSRLQVGVGTIPVVSCVTGTWLTAKQLADPRYHARHAREPVRFAAGLATLTGEEPFLM
5 LEVGPGSTLAALAREHSNARLPVVTSLRHARQATPDQYLLLETGLCLRHWGVSVDWGAHAGRS
10 RRLVSLPGYPFSGAVRRLAGDPLRLLAGARAVAAPSGTRQLSADARDLPNTPEPTSGAVSAIK
APIAAADPGLYRLSWRQAGTAPLGPDLGPPRDWIVFASDSHLLQALRANLGTRAQRVTLVTP
GOEYAAEPSGFRLRPDQIDDYRALWADLAQTGIVPRYIAFLAPFMYRARMAGDASTLDEVREG
GFLPLTRLIQTRPPGGPSGLLSLTIVTPAALALGDEATRPEWAILHGMVAGLSRDYPEWRVVS
15 IDGGDPSPHRCEGLARLIALHAVDEAGPTRLALRGLHAWVPQCEHVQPATIPGAGMWREGGVY
MITGGFGGIGLALARALAREARAKLILVGRNLPTAPIDLEAWDAPPLILTADVADDEAMRRVF
DAAHARFGAIDGILHAAGVPGGSLFANQSDAAFEDVLHAKVRGTLVLQGLRAIDAPLLLMSL
20 DAWLPGPQGTAYAAANAFDLAFASLRRREGEVPVYSGWDSWCEVGMAARVAARSADERGRLAR
EGISPRQGWQALSRLALDPPHLMISRTDLTSRWHSRSSPTPVASSEPEVALPRWTASACQAV
25 IERVWCEHFATAAVPPDGNFFELGASSFDIVQLSARLQQQFGRDVSHTVLYSHPTVALLAGYF
15 ANDPTPSGAAADERDEAVRRGRDLLKSRRRGV

30

pEPOcos6_ORF12 sequences:

35

20 (1) nucleotide sequence

Seq ID No 63 (>pEPOcos6_ORF12.seq)

40

ATGACCGTGGAGCACGAAACCGGATTGAAATCGCCGTCATCGGGCTGGCTTGCCGCGTTCCC
GGCGCTGCCGACGTGGCCGCCTTCTGGCGCAACCTGGTCGAGGCCAAGGAGAGCGTGCGCTTC
25 TTCGAGGACCACGAGCTGCGGGCCCGCGCGTCCCGAGGAGATCTTGCGCCTGCCCAACTAC
GTGAAGGCCAAGCCACTGCTCGCTGATGGCGAAGCTTTCGACGCGGACTTCTTCGGGTTCCAT
CCGCGCGAGGCCGCTACCTGGACCCGCAAGTTCGGCTCCTGCACGAATGTTGTTGGACCGCG
CTGGAGGATGCCGGCTACGATCCCGCGCAGTACGCCTACCCGATCGGGTTGTTCCGGGGCGTC
45 TCCAGCAATCTCTCGTTCTGTTCGACCGCATCGATCCGCGCGACTCCCCCTGCAGAAGCGC
TATGTGGCCGAGCTGAACGCGGCCTCCTTCGCCACCCAGATCGCCTACCGGCTCGATCTGAAG
30 GGGCCGGCCATTTCGATTCAAACCGCCTGTTTCGACGTCACTGGTGGCGATTACCTGGCGGCG

50

55

5

130

10

5

15

20

10

25

15

30

35

40

25

45

30

50

55

CAAAGCCTGATCGGCGGCGAGTGCCACATGGCCTTGGCCGGCGGAGCGACCTTGGAGGTCCCC
AAAAAGCCCCGGCTATCTCTACCGCGAAGGCTACATCAACTCGCCGGACGGCCACTGCCGGGCC
TTCGACGCCGACGCGGCCGGCACCATCTTCGGCGACGGCGTCGGCATCGTCCTGCTCAAACGC
TACCGCGACGCCCTACGCGACGGCGATCACGTGTACGCAGTGATCAAAGGCTCGGGCATCAAC
AGTGACGGCCATCGCAAGGTGTCTTACACGGCGCCGGGCAAGAGCGGTCAAGTGGCGGTGATC
CGCGCTGCCTGGCGGCGGCCAGGTAGAGCCGCAAACCATTCGCTTCGTCGAGGCCCCACGGG
ACCGGCACACTCGCCGGCGATCCGATCGAGGTAGAGGCGTTGACGGAGGTCTTTGCCGAAGCG
GTCGCGGTACCTGCGCCCTGGGTTCGCTGAAGACCAACATCGGCCACTTGGATGTGGCGGCG
GGCGTGGCCGGTTTCATCAAGGCGGTCTTGGCGCTCGAGCGGCGCTCCTCCCGCCAGCCTT
CACTTCGTCCGGCCCAACCCGGCCATCGATTTCAACGGGCCCTTCTACGTTTGTTCGCCAAATC
GAGCGGTTGACGGAGAACGGGCGGTTGCGGGCCGGGGTGAGTTCCTTTGGCATTGGCGGCACC
AATGCCACCGTGATTCTGGAGGAACCGCCGGCGCGGAGGCGAGACTGCCGGCCGGGAGCCCCG
CCAGGCGCGAGTCCGTTCTGTTCCTGCTATCGGCCAAGACGCCGATGCGCTGGCAGGCCGT
TGCCACGACCTTGCCGACCACCTGCGGGCGCACCCCGAGCTCCTCCTGGCCGATGTGGCCCTC
ACTCTGCGAGATGGGGCGGGCGTCTTCGCTACCGCCATGTGGTCCAGGCTGCGACGGCGGAG
GAGCTGATTGCGGCTCTGGGAGCGTTCCGACAGGAGTCCATCCGCAAGAGGCGGAATCGAGTA
CAATGGGTGTTGGCAGGCGAGGCGATGTGCTTGACGCCGGTTTGGCGCTGTACGCCGATTGG
CCGGTCTATCGGGAGCGGGTCGACGTCTGTCTGGCGATCGTCGCCAAGCTGCGCCAAATCGAC
GGCCGGTCATTCTACATGAGTGAATCGAGCGACCGCGCGAGGTTCTGCGGAATGGTCGACG
GCGCTGGCGTTTCATGTTCCACTGCGCGCTGGCGCAAGCCCTGAGCCAGGCCGGCCTGCACCCG
CAGCGCATGTGGAGCCGTGGGCTGGGCGGACAGGTCGGCGTGTTTGGCCGAATCCCTGTCTG
TTGGAACAAGCGCTGGCGCTGGTGTTGTGCCAGACACCGGTTCCCGGCGATGCCACACCTCAG
CGCGAACGCTTGGTTCGGACACTGGAAGGCTGCCGGTTTCGTCCACCACGATTTTGTATTCG
GCAGACAGCTCGGGTCGACCCCTGGACCTCGCCGAATTCGCTCATGTGATTTTTGGTGCGGT
GGCCAAAGCGCCTCGCCCAATGAGGCGGAGCTGCGCTCATGGAGCGACGCCGCGCCCGAGCTG
GTGACCTTGGCGATCGGCCCATCTTTCTCGAGGCCGCTCCGGGACGGTGGGTCTGGCGATC
GACCCCAAGCGACCGATGACCTGTGTTTCAGCGCACGGTGGCCGCGTTGTGGGAATGGGGATGT
GACGTGCGCTGGGCTGCGTTACCTCGTCGACCGGGCGTCGGGTTCCCTGCCTACCTATCCC
TTCGTGCGGGTAATTCCCACGATCGGCGACCCCTTCGCGGAGCAGGCGCGGAGGATGACTTG
ATTGCGGCGAGCGCTTCCGCGTCGSCCGGATCGCCGCCGAGCCGTCCGCAAACCTCGGCAGCG

5

131

10

15

20

25

30

35

40

45

50

55

GAACGCCCACGCGCCAGTCAAGCATCGCCTCGGCAACCACACCGGCTCCGTCTCATACGTG
GCCAGCGTGGCGTGGCCACCATTCTCGAAACCGTCCGTGCCTATTTGCGGTTCGCCGCCGTG
CGTTCCACCGACGCCTTCTTCGAATTGGGCGCGTCCTCGCTGGATTTGGTCAACCTGGGCCAG
CTCCTTTCCGATCGTCTCGGCCGCGAGGTTCCGACCCTGCTCCTCTACGACCACCCAACACCG
5 GACCAGTTGGCGCTGGCCCTGACATCCGCGGCGCTCAGCGCAGAGGCGCCGCCCTTAAGGGGC
GGTCATCGCGCATCGACTTCCGGCACAGCCGCGAGCTCGGCCGCTCCACCGCACCGACGTTT
15 CCGGGGACGCTCACTCGAGCCAGCTTCGTTTCGCGAGCAGGACATCGCCATCATCGGGATG
GCCTTCCGGGACCGGGCGCCGACGACCTGGACGCGTTCTGGAACAACCTGGTGAAGGGGTC
GAGTCGATCACCTTCTTCAGCGAGGACGAGCTGCTGGCGGGCGGCGTCCCCCGGAACATCTG
20 GCCTCGACGCGCTACGTGCGGGCCAAGGGGAACTGACTGGGATGATGGATTTGAACCGGAA
TTTTTCGGTTATTTCGGCGCGGAGCGGCGGTTCATGGACCCGAGTTCGCGTGTTCACCGAA
TGCTCTGGCACGCACTGGAGCACGGCGGTACGATCCGACCCGATGCGCGGCATCGATTGGC
25 GTCTACCGCGCGTGACCAACCACCTGCCTTGGCTGATGCGAACTTTGCCGCACCTGACCGAG
GAGGAGCAATTTCGGCGCGCTGCTCCTCACCGACCGCGAGTTTTTCGCACCGCTGCTCTCCTAC
15 AAGGTGGCCCTGCGCGGACCCGCTATTTTCGCTGCAAACCGCCTGTTTCGACGTCGTTGGTGGCG
ATCGGCACGGCCTGTCGGAATTGCGCGCGGGTGCCTGTCAGATGGCCCTAGCGGGCGGCGTG
30 ACGGCCAGCATCGAGCGCTGCGGCTACTTCCACCAAGAAGGCTACATCCTCTCGCCTGACGGC
CACACGCGCAGCTTCGACGCGGCGGCCCGGCGCACGGTCTTCGGCGACGGAGTCGGCATGGTG
CTGTGAAGCCGCTGGCCCAAGCCTTGGCCGACGGCGACACGATCCACGGGTGATCAAGGGA
20 ATCGGCATCAACAACGACGGCGCGCGCAAGGTTCGGCTTACCGCACCTAGCCGGGCCGGTCAG
ACCGAGGCGATTTCGGCCCGCGCTGCGCGACGCGGGGTGGCGTCGAACCGCGTCAGCTACGTG
35 GAGGCGCATGGAACCGCGACAGAAATGGGCGACCCGATCGAGGTCGAGGCCTTGACCCAAGCC
TTTCGCGCCGAAGCCGACGGTCCGCTTCCGCCCCGCTCCTGCCTACTCGGCTCGGTGAAGTCC
40 AACGTGGGCCACCTGAACGCCCGCGGCGGCGTGGTGGTCTGGTAAAAACCGTGCTGGCGCTC
25 CAACACCGCCGCTGCCGACCAGCCTGTTCTACAGTCGCCCAATCCACACATCGACTTTGCG
GCGAGTCCGTTCCGCGTGAACGGCCAGACTTCGGATTGGGTGCGGCCAGAGGGGACGCGGTTG
CTGGCGGGAGTGAGTTGTTTCGGTATCGGGGAACCAACGCCACCTGATCGTCGAGGAGGCG
45 CCGAAAGCGCTACCGACGACAGCGGCACCTCTGTGACGGAGCCGAATGACCTCGACGCGGGC
GACGCCGACGGGCTAGTGCTGCCGATCTCGGCCCGCACGCCGACCGCCCTGGCGCACATCGCG
30 ACCAACCTCGCCAATCACCTGGAACGACATCCGACCATCGCCCTGGCCGACGTCGCCCTGACC

5

132

10

15

20

25

30

35

40

45

50

55

CTTCAGCTGGGCGTCGCCAATGGCCCCATCGCCACAGCCTGATCTGCCGGAATCGAACGGAG
GCGATCAAGCTGCTGCGCGCCGTCGTCCACTCCGCGGAGGTGCCGCCAGCTCAGGCGCCGGTC
TCGGATGCGCCGCGCTGTGTTTTCTTTTCCCGGCCAGGGCGCCCAATACCCGAGCATGGCC
CGCGACCTGGTTCGAACTGTCCCGACTTCGCCCTGCACCTGGACCCCTGCCTCGACCACTTG
5 GCCGAAGTGTTCCTCGAAGATCCGCGTTGCATCCTGTTTCGGCGATGGCCCCGCCGATCGGCTC
15 GACCAGACGGCCTACACTCAGCCGCTGCTCTTCTCCGTGTCCTACGCCTTGGCGCGCTGTTG
GGCGATTTCCGGATTTCGCCCCGATGCGATGATCGGCCACAGCCTGGGCGAATACGTGGCGGCC
TGCTTGGCCGGGCTTTTCTCGCTGAGCGATGCCCTGCTGCTGGTGAGTGAACGCGGCCGCTG
ATGGGCTCGGCCGCGCGGAGCGATGTGGCCGTCCCTTGCCCGAATGGGAAGTGGAGGAA
20 CGCCTGGAGCTTCTGGCCGACGACCGAATCAGCATCGCGCGGTCAACACCGCCGAGAGCTGC
10 GTCATCGCGGGACCCAGCGAGGCGATCGAGCGCTGCGCCCAGCGCTGGGCCGCGCAAGCCCTG
ACCTGTACGCCGCTGCGCACGTCCCACGCCCTTCCACTCCGCGATGATGGAGCCGATTGTCGAA
CCCTTCGSCCATGTCTTGGCACGGGTACCTTCGCGCGCGCCGCGCGCTGGATCTCGAAC
25 CTCGACGGCAAGCCGATCGATTCCGCGCGGTGATGCAGCCGACTATTGGGTGCGCCACCTG
15 CGCCAACCGGTCCGCTTTCACGAGGAGTACGTACCTGTTGGCCGAGGACACCCATGCTTGG
GTGCAAGTGGGTCCCGGCCGAACCTGTCTCTTCTGTCGCGCGCCACCCGGCTACCGTCAC
30 CAGCCAATCGTCAACCCCATGCGCCATGCAGTCGAGTCGACGGGCGACGTGCGCCGGTGGCGC
CAAGCGCTGGGCGAACTATGGCGGGCCGGCATGCCGGTCCCTGGGAGCGGCAGCGCGCGGC
CGGCATGCCGGACGACGTGTGCCGCTGCCGGGCTACCCCTTCGAGCGCGGCCCTTCGCGGCC
20 CGAAGACCGGTGGAGCTGGCGCAGCCCGCGCCCAAGGCGGAGCTGGTGA AAAAACCCGATCCC
35 GCGCGGTGGCTGTACCGCCGCGTCTGGCGCCCTGCCAGGCTGCGGCCGGCGGACTGGCGGTG
CAGGCGACCGTTCTGGTCTTCGGCGACGGGTCCGAGCTGTGCCGCGCGGCGGTCCGCTCAGGTG
CAGCGCCAGGGGCTGAAGTGCCTCTCGATCACCGCGGGCCGCAATTCGCGCGGAGAGCGAC
40 ATGCGCTTACGCTTGACCCCGCTGATCCGCGCCAGCTCGACCAGCTCTTCGCGGCCCTCGAT
25 GGCTCAGGCTCGCGCCGCGGTACGTCTGCACCTGCTGACCTGAACCCGCCCCCGGATGCC
TCGSCGATCATCGCTCACAGCTACTACAGCCCGATGGCCTTGGCTCATGCCTTGGGCGCCAC
GAGATCGCGCCTGTCTCGATCACCGTCGTACCGCCGGGTGCTCGCCGTGCGGAAGAAAGCG
45 ATTCGCGAGCCGCTGCAGGCGCTGATCGTGGGCGCGTGCCTGGTTCATCCCGAGGAGTTCCC
GGCTCAGCGTTCCGCTGCTGGACGTCAACGTCGACGATCCGGCACCGCGTCTGGCGGAGCGG
30 CTCGTGGCCGAGCTCTCGGGCACGATCACATGGTGGCGCTGCGCGCGGCGAGCGCCTAGTG

5

10

15

20

25

30

35

40

45

50

55

GCCGATGTCGATCAAGTCGATGGCCTCGGTGTGGGGATCGCCAAGGTGCCCTTGCGCCGCGAG
GGCCACTACCTGATTCTCGGCGGCCTGGGCGATATCGGCTACCACTGTGCCCCGCTATCTGGCC
CAAACCTACCGCGCCAAGCTGACGCTGACCGCGCGTTTCGTCACTCCCGCCGCGCGCTCGTGG
GAGCGAATGCTGCGCGAGGGAAACCTGGATTCCCGGCAGCGCACGCGCATCGAGCGCGTGTG
5 TCGCTAGAGGCGTGCGGGGCCGAAGTCCAGACGGCTGCGGTGCACTTGGGCGATCGCCATCGC
15 TTGGCCGATGTGTTCCGCGAAGCACGGGGCCGATTTCGGCGCCATCGCGGGCGTGATTCACTCG
GCGGGGATTCCGGGACACGTCCACTCGATCGACGAGCTGGTGCGCGTCCGCGACGAAGCCCAA
TTCACCGCGAAGGTTTCGAGGGCTGCACCACCTGGCCGAGGTCGTGATCCGCTGAACCTCGAC
TTTTGTCTGCTGTTCTCCTCGCTCTCGACCGTCTTCGGCGGGCTCGGCTACGGCGCCTATGCA
20 GCGGCCAACGCCTACATGGACAGCTTCGCCCGCCGCCACGATCGGCCGGACGAATGTCGTTGG
10 ATCGCGGTCAACTGGGACGCCTGGCTGTTTGAAGCCAAGACGTCGTCGGTCGGCGCCGAATTG
GCGCGCCTGGCGATCGTGCCCGAGGACGCTCCGGCCCTGTTGCGCGGGTGCTAGAGCGACTT
CCGCAATCGTTTCATCGTGTCCACCGCCGACCTTCGGGCCCGCATCGACACTTGGATCCGGGAC
25 AAGAACCGCGTCCCGCCCGCCGAGATCCGAGCGGTTCAACCGCGACCGGACCTGAGCCAGGCG
15 TACGCCCCCGCGATCGGCCCGCTGGAGATTCAACTCTGCGGGCTGGTCTCCGCCTATTGCCGG
TTCGACCGGATCGGGCGGGACGATTCTTCTTCAAATCGGCCTCAGCTCGTTGACTTGATC
30 CAGCTCAGCTCGCGCATTACCGCATCACCGGCAAGGATCTCAATACGACCCAACTGTTTACG
TACCCACCGTGCGGCCTTGGCGCTCTTCTTCGGCGGCGAACCGGAGGGGCTCGCGGCGGAG
GAGCCCGCCATGGAGAACCTGTGGCTGCAACGAAGCGATGCGACCCCTCGATGAG

20

(2) peptide sequence

Seq ID No 64 (>pEPocos6_ORF12.pep)

MTVEHETGFEIAVIGLACRVPGAADVAAFWRNLVEAKESVRFFEDHELRAAGVPPEILRLPNY
VKAKPLLADGEAFDADFFGFHPREAAAYLDPQVRLLECCWTALEDAGYDPAQYAYPIGLFAGV
40 SSNLSFLFDRIDPRDSPLQKRYVAELNAASFATQIAYRLDLKGPAISIQTACSTSLVAIHLAA
25 QSLIGGECHMALAGGATLEVPPKPGYLYREGYINSPDGHCRADFADAAGTIFGDGVGIVLLKR
YRDALRDGDHVYAVIKGSAINSDGHRKVSYPGKSGQVAVIRAALAAQVEPQTI RFVEAHG
45 TGTLAGDPIEVEALTEVF AEAGRCTCALGSVKTNIGHLDVAAGVAGFIKAVLALERRVLPPSL
HFVRPNPAIDFNGPFYVCRQIERLTENGR LRAGVSSFGIGGTNAHVILEEAPAPPEARLPAGSP
30 PGASPFLEPLSAKTPDALAGRCHDLADHLRAHPELLLADVALTLQMG RASFAYRHVVQAATAE

5

10

15

20

25

30

35

40

45

50

ELIRGLGAFRQESIRKRRNRVQWVLAGEAMSLDAGLRLYADWPVYRERVDVCLAIIVAKLRQID
GRSFLHEWIERPREVPAEWSTALAFMFHCALAQALSQAGLHPQRMWSRGLGGQVGVVLAESLS
LEQALALVLCQTPVPGDATPQRELRVLTLEGCRFRPPRFLISADSSGRPLDLAEFAHVDFWCG
GQSASPNEAELRSWSDAAPELVTLAIGPSFLEAASGTVGLAIDPKRPMTCVQRTVAALWEWGC
5 DVRWAAFTSSTGRRVPLPTYPFVRVIPTIGDPLRGAGAEDDLIAASASASAGSPPEPSANSAA
10 ERPRAQSSIASATTPAPSHTSASVAVATILETVRAYFGFAAVRSTDAFFELGASSLDLVNLGQ
LLSDRLGREVPTLLLYDHPTPDQLALALTSAAALSAEAPPLRGGHRASTSGTAASSAASTPTF
PGDAHSQPSFVREQDIAIIGMAFRQPGADDLDAFWNNLVEGVESITFFSEDELLAAGVPREHL
ASTRYVRKAGELTGMMDFEPEFFGYSAREAAVMDPQFRVFHECSWHALEHGGYDPTRCAASIG
20 VYAGVTNHLPLWMLRTLPHLTEEBQFALLLTDREFFAPLLSYKVGLRGPALSLQTACSTSLVA
10 IGTACREL RAGACQMALAGGVTAIERCGYFHQEGYILSPDGHTRSFDAAGTVFGDGVGMV
LLKPLAQALADGDTIHAVIKGIGINNDGARKVGFTAPSRAGQTEAIRAALRDAGVASNRVSYV
EAHGTATRMGDPIEVEALTQAFRAEADGPLPPGSCLLGSVKSNGHLNAAACVAGLVKTVLAL
25 QHRLPTSLFYQSPNPHIDFAASPFVRNGQTSDWVAPEGTRLLAGVSSFGIGGTNAHLIVEEA
15 PKALPTTAAPLSTEPNDLDAGDADGLVLPISARTPTALAHATNLNHLERHPTIALADVALT
LQLGRRQWPHRSLICRNRTEAIKLLRAVHSAEVPPAQAPVSDAPRCVFLFPQGGAQYPSMA
RDLVRNCPDFALHLDPCLDQLAELLPEDPICILFGDGPADRLDQTAYTQPLLSVSYSALARWL
30 GDFGIRPDAMIGHSLGEYVAACLAGLFSLSDALLLVSEGRMLGMSAARGAMLAVPLPEWELEE
RLELLADDRISIAAVNTAESCUIAGPSEAIERCAQRWAAQGLTCTPLRTSHAFHSAMMEPIVE
20 PFGHVLARVTFAPPRARWISNLDGKPIDSAAVMQPDYWVRHLRQPVRFHEGLSHLLAEDTHAW
35 VEVGPGRTLSSFVRRHPAYRHQPIWNPMRHVESTGDVRRWRQALGELWRAGMPVAVERQRRG
RHAGRRVPLPGYPFERRPFAARRFVELAQAPKAELVKNPDPARWLYRRVWRPAQAAAGGLAV
QATVLVFGDGSEL CRAAVAQVQRQGLKCVSITAGRQFARESDMRFTLDPADPRQLDQLFAALD
40 GSGSRPRYVLHLLTLNPPPDASAIIAHSYYSMALAHALGAHEIAPVSI TVVTAGVVAVADEA
25 IREPLQALIVGPCLVIPQEFPGLSWRLLDVNVDPPAPRLAERLVAELSGTDHMLRGGERLV
ADVDQVDGLGVGIAKVPLRREGHYLLILGGLGDIGYHCARYLAQTYRAKLTLTARSSLPPRASW
ERMLREGNLD SRQRTRIERVLSEACGAEVQTAAVDLGDRHRLADV FREARGRF GAIAGVIHS
45 AGIPGHVHSIDELVRVRDEAQFTAKVRGLHHLAEVVDPLNLD FCLLFSSLSTVLGGLGYGAYA
AAKAYMDSFARRHDPDECRWIAVXWDWLFEAKTSSVGAELARLAIVPEDAPALFARVLERL
30 PQSFIVSTADLRARIDTWIRDKNRWPPAETRAVQPRPDL SQAYAPPIGPLEIQLCGLVSAYCR

55

5

135

10

FDRIGRDDSFFEIGLSSFDLIQLSSRIHRITG:DLNTTQLFSYPTVRALALFLGGEPEGLAAE
EPAMENLWLQRSDATLDE

pEPOcos6_ORF13 sequences:

5

15

(1) nucleotide sequence

Seq ID No 65 (>pEPOcos6_ORF13.seq)

20

10

25

15

30

35

40

25

45

30

50

55

ATGAAATACGAAACCACCGGATTGGAATTGGCCGTCATCGGTCTCGCTTGCCGCTTTCCAGGC
TCACCCGATCCCGAACAGTTCCTGGTTCGAATCTGCGCGCAGGTTCGCTCCGGAATCCGCCATTTC
AGCGATGCCGAGCTGAGCCACATCCCCGCATCCCTGCGTCACCATCCGCATTACGTCAAGGCC
AAAGGCGCGCTGGACCACGCCGATTTCCAACAGCCTTCTTCGGCTACTCGCCCAAAGAGGCC
GAGGTGATGGACCCTCAATTCCGGCTGCTCCATGAGTGTGCTGGGAGGCGCTGGAGTCAGGC
GGCTATGCGCCGAGCCAATTGCGGGTTCGGATCGGCTTGTTCGCGGGCGCGCCTTCAACGAC
GGATGGATCGCCGGTACCTTCGACCGGCTGCGCACCGGCGTGGGTTTGTAGCTCCCTGGAAACC
GCCTTCTTGACCCTGCGCGATTACCTGACCACCCAGATCTCCTATCGGCTCGATCTGCGGGGC
CCCAGCCTGCTTGTCCAAACCGCCTGCTCGTTCGTCGCTGGTGGCGGTCCAGCTCGCCACGAG
GCCTGATCTCCGGCAATGCGCCCTGGCCTTGGCTGGCGGCGTGTGCGGACCGATCCGCTG
CATTCCGGATACCTCTATGAACCCGGCAACATCTACGCGCGGACGGCGTCTGCCGACCGTTC
GACGAGGCAGGCGCCGGTACGGTCTTCGGCGACGGGTGCGGCATGGTCCTGCTCAAGCGGCTG
AGCGACGCCCAGCGGACGGCGATACGATCTGGGCGGTATTTCGCGGGCGGGCGTGAACAAC
GACGGGCACCACAAGGTTGGCTACACGCTCCTGGCACGAGGGGCCAGGTGGCTTTGCTTAAA
AGTGTATTATCGCGGAGCCGGGTGACCCGGCGACGCTCGGCTACCTGGAGGCCCATGGCACC
GGCACCGCGCTCGGCGATCCAATCGAGGTGAGGCGCTTACCCAGGCCTTCGCCAGCAAACGT
CGCGGCACCTGCGGCTTGGGCTCGGTCAAGGGCAACCTGGGTACCTCAACACGGCGGCCGGC
ATCGCTGGACTGATCAAGGTGGTGTGCGCTGAAACATCGCGAAGTGCCACCCACCCTCAAT
CTGCGCGTCCCAATCCGAAAATCCGCTTCGACGAGACGCCGTTTTTCCCAGTCGTCGAGTTG
CAACCCTGGCCAAGCGGGACCGGCCCTTTCGAGCCGCGTGAGCTCCTTCGGCATCGGCGGT
ACGAACGCCCACGTCATCTTCGAGGAGGCACCGCCGACGGCCAACCCGGCGCCACACGGCAGA
TTCCGACTGTTGCCGCTTTCGGCCAAGACACCGGCTGCGCTCGAAGCGAAGCGCCGCGATCTG
GCCGCTTCCTCGAACGCCACCCGGAGACCTCCTTGGCCGACCTCGCCTTTACCTGCAACGC

5

136

10

15

20

25

30

35

40

45

50

55

GGCCGCGAGGTCTTCAGTCACCGCGCCTGCCTCGCCGTGGAGACCTTAACGTCCGCGCGCACG
CGGCTGAGCGGCGAGTCGTGAGCACTTGCGTGGTGGGCCCCGCGCCCAGCGCCATATTTCTG
TTCCCTGGTCAAGGCAGCCAGCTCGCCGGGATGGGCCGCGGTCTGTATCACCATTTTCGAGCCG
TTCCGCACGGCCGTGATGCCGTGTCTGCGCGAGCTGGAGCCAGGACTGCGGCAAGCGCTCAGC
5 GCCCATTTCGATCCGAATCGCGGCGCGGACCCACCCGATTGACGACCTTCGTCCAACCCCTTG
15 TTGTTCCCTCGTCGAGTACGGGGTGACCGAGTGGCTACGCTGCTTGGGTGTGCGGCCAACAAATG
GTGTTGGGTACAGCTCTGGCGAGTATGCCGACGCTGCGTCGCGGGCGTTCTGTGCGCCGTCC
GCGGCGGTCTCGCTGCTGGCCGAGCGCGAGCGGCTGCTGCGCGACCTGCCAGCCGGCGCCATG
CTCGGCGTCCCGCTGGCCGCCGAGGCGCTCGAGGCGATGTTGCCCAGCGCTCTCGATCTGGCG
20 GCGATCAACGGCTGTGAGCTTTGCGCCGTGTCGCGGCGGTGCGGCGGTCCACGCCTTCAAG
10 GCCCAACTGGAAGCCGCCGACATCACGCCCGCTGTTGCACACCGATCGCGCCTTCCACTCG
CGGCTGGTAGCACCGGTGCTTGACCGGTTCCAGGCAGCCGTTCAACACGTGGAGCTGCGGCGG
CCGCAAGTACCTTACCTCTCGACCGTCAGCGGCGATTGGAGGCGGATGGGCCGGCGAACCCG
25 CACTACTGGGTGCGTCACCTGCGCGACACGGTGCGGTTTGGTCCAGCCCTGGAGGCGCTGCCG
15 CCGGTGGATTCCCTTCGTGTGCATCGAGGTGGGACCAGGCTCGGCCTTGAGCACCATGGCGCGC
GAAACGTGGGTTCCCAGGCGCGACTGATTTCGTTGCTGCCGCGGCCGGAACGGGGCAAATC
30 GAGCCCGGTCCGGTATTGGAACGACTGGCGGCGCTTTGGCGCAGCGGGTTGACATTGGATTGG
TCTAAATTGACGGGCGGCGAAGAGGGTCATCGAATCCCTTGCCAGTCTACCCGTTTCAGCGC
AGCCATCTGTGAGCTCCCTGGCGGCGGGCCACACGCCTTCGTGCGGCGCTGCAGTCGAATCA
20 GGCGCCATCCTTGCCGAGCGATCCGACGGGAAAACGCTGAAACCCGGGATTGCCCGCTGCCA
35 ACCGCCACGCTCGAGCCCAAGGCGGTGCTCCGGCCCCACTCGAGGCTACCGACGCCGAGGT
ACTGCGAGCGACTGGCCGAACTTTGGCGCGAGTTGCTAGGGTTGACCTCGATTGGGCCCCGAC
GACCATTTCTTCGACCTGGGCGGCCACTCGCTGACCGCCACGCGGCTGCGCGCCCTGATTAC
40 CAGCGGTTTCGATGTGATCTCGGGCTCGACGAAATCTTCGCTCATTGCGGTCTCTCCAGCTG
25 GCCGCCCCGTATCGAGGCGGCGGCCAAGAGCCGATTTTCTCCATTCCAGCGCGCCGGACCAG
GAC3ACTATCCCTTGTCATCCGCCCAGCAGCGGATTCACAGCATCGTCACGAGGGCCGAGGTC
GGCACTGCTTATAATTTCCGATCGTCCTCGAGCTGCAGGGCGCTCTGGATCGAGTGCGATTTC
45 GAGCGACGTTGCGGGCATTTGTTCCGGCGTCATGAGGGGTTCCGCACCCGCTTTGTGATGCGC
GATGGCGGGCCGCGCCAGCGCATTGTACCGGACGTGGCGTTTCGCCTGCCGCTACCCAGGTC
30 GAGCCAGAGCAGGTTCCCGGGCGCATCGAGGCCTTCATCCGTCCCTTCGATTGGAACGCGCG

5

137

10

15

20

25

30

35

40

45

50

55

CCGCTGTTCCGCGCGGAGCTGTTGCAGTTGGCCGAGCAGCGCCATCTGCTACTTTTCGACATG
CACAACCTTAATTGCCGACGGTATCTCGCTCAACCTGTTTCGTGCGCGATTTCGCGGCCCTGTAC
CATGGTCGTCCGCTGGCGCCGCTGAAACTCCGCTATCGCGACTATGCCGTTTGGCAAGAGGCG
CGGCTGGCCTCCGATGACCTGCGCA3CCAGCGGAATGGTGGCACC GGCGGCTTTCGCCGCCG
5 GTCGCCACGCTGGCGCTCCCTCCCGATTTCGCGCTCCGGCGGTGCGCCGCTACAAGGGCCGT
AATGTGGTGTTCACCTGGACCGGGAGATCCGCGACCGCTGGTGGCCCTGGCTCGAACCAG
GGGGTCACCATGAACGTGATGATGCTGGCGCTCTGGGCTGCGCTGCTGCATCGCGAAACCGGC
CAATCGGAGCTGGTGGTCGGATCGCTGCTCGGCGGGCGGCCGACAGCGAGCTGCATCCCGTG
ATCGGGCTCTTCACCAACTTTTGCCCTTGCGGTTGGCGGTGAGGGATCGACCCGCTTCGAT
10 CGCTTCCTTGCCGCTTGCCACCAGGTGTTTCTCGAAGCCTATCAGCGCCAGGACTATCCGTTT
CACTTGTTAGTCCAGGAATCGTGCCGGTCAGGGACCCGTCGCGGTGCGCGCTGTTCCAGACC
TCGCTCGTCTACCACAACGAAATTGACGGCAAGACCAAGCTGGAATTGGAAGGGCTGAAAGTC
GAAGTGTTCCCTTCGAAAAGGGTGTGGCGAGGCTGGATTGGAAGCTGGATGTGACACCTTTT
TCCGACCGACTCGAATGTGTTTTGCAATACGACTTGGATCTGTTCTGCGAGGAGACGATGCGC
15 GGCCTGATCGCGCGGTTCCAGGCGTTGGTGGCGGGGCTTGTCGCCGATCCGGCGCAATCGCTC
GCCGCCGCGAGCGTTTCCGGGAAGCGGGCGCTGCGCGCGGGCGTGCCACGGCAAGCGAATCG
TCGCCGCGAGTCACTGCCCGCAACCATCGACGGCGTACGCCACTCCCTCACCAGTCACCG
TCGCCGGTAGTCCTGACGGGACCCGCCGACCTGCCCGGATCTTGGCGGCCTACGTGGGGCAG
AACCCCATCCGTTCCGCGATCCATCGGGGTCTCATTTTGGAGGCGCGCTGGGGTTGCGAGCG
20 CTGCGGTGCGCGCTGGACGCGAGTGTCTGGAGAACACACCCATTGGCGCAGCGTGCCTGCGGGC
GATCGCGCGCGGCGGTGGATAAGTTGGAATTGACCAGCCTGGTGGCGCTCGACGACCTGCGC
GGGTGGTCAATCCTCAGGCGAATGCCTTACCCCTGGCTTGGCGCGATCTGGCGATGCCGTTT
GGGGAGGGGCGTCCCTGTGGCGACTCCGCCTGGCGTGGTGGCTCCATCGCGCTGGTTGCTA
40 TTGCTGACGGTTCATCCATTGATCGGCGACAACGGCACGGTCGACCTCTTTCTGGCGGCACTC
25 GCCGATCACCTGCGCGCGCGTCCGCTTTTCCCGTAGCACCGCTCGATGAGGCCGAGCTGGAG
GCCGAGCTGAAGTGGGGAGAGGAAGGGGAGGGCTCGGGCTGACCGGATCGCGCCGGTCTCTG
GGCCAATTGCGCGAAAGTCGGCTGAGTCTGTGGCCAGATGTGGCTGGACGAGGTCTGTGCG
45 CGCCACGACCTACCCCGCTAGAGGTCTTGGCGGCCCGGCTCCTCGATTGGACACGAAGCCAC
GGTCACGGGTGATCGCTTTGTGGACGCCGCTGCCCGAGGACCATCCGCTTCGCGATGAAGGC
30 CGCTGCCTCCAGGTTGCGCTGCTGGAGGGGCGCGCGTGCAGCGAGGAGCGGGCGATCCAAGC

5

10

TGGCTCGAGCAAATCGCCTTGAGACGGGGTACCCCTGCAACGGAGGTCGTTTGCCTACTCCG
ACCCAACGGGCAGCCATCGACCTCGCGCTGGCCTGGCTGCCGAGCCGCCTCTTCACGGTTTG
GTCCGAACCGTTTCAGCCGTGGCCGGAATCTCCATTGGTCTGTCCGTTTCCCCTCAATCTCGCG
TTCCGGCCAAGCCATCCAATTGCCTACGCGCTCAAGCACGAGGCCACGCTCGCGGTCACGGCA
5 CGGGCGCGCGATCTGATGCGTTTCTCTGACGGCTTGGGCCCCGAAAGC

15

(2) peptide sequence

Seq ID No 66 (>pEPOcos6_ORF13.pep)

20

10

25

15

30

35

40

25

45

30

50

55

MKYETTGLELAVIGLACRFPSPDPEQFWSNLRAGRSGIRHFSDAELSHIPASLRHHPHYVKA
KGALDHADFEPAFFGYSPKEAEVMDPQFRLLHECCWEALES GGYPQS QFAGRIGLF AAAAFND
GWIAGTLDRLRTGVGLSSLETAFLTLRDYLTQISYRLDLRGPSLLVQTACSSSLVAVQLAQQ
ALISGECALALAGGVCATDPLHSGYLYEPGNIYARDGVC RPFDEAGAGTVFGDGCGMVLLKRL
SDAQRDGD TIWAVIRGAGVNNDGHHKVGYTAPGTRGQVALLKSVYRASRVDPATLGYLEAHGT
GTALGDPIEVEALTQAFASKRRGTCGLGSVKGNLGH LNTAAGIAGLIKVV LALKHREVPPTLN
LRRPNPKIRFDETPFFPVVELQPWPSGTGPLRAGVSSFGIGGTNAHVILEEAPPTANPAPHGR
FRLLPLSAKTPAALEAKRRDLAGFLERHPETSLADLAFTLQRGREVFSHRACLAVETLTSART
RLSGESSSTCVVGPAPSAIFLFPQGSQLAGMGRGLYHHFEPFRTAVDACLRELEPGLRQALS
AHFDPNRGADPPDSTTFVQPLLFLVEYGVTEWLRCLGVRPTMVLGHSSGEYAAACVAGVLSPS
AAVSLLAERERLLRDL PAGAMLGVP LAAEAL EAMLPDALDLAAINGCQLCAVSGPVAAVHAFK
AQLEAAGHHARLLHTDRAFHSRLVAPVLD R FQA AVQHVELRRPQVPYLS TVSGRLEADGPANP
HYWVRHLRDTVRFGPALEALPPVDSFVCIEVGPGSALSTMARETLGSQARLISLLPRPRGTQOI
EPGPVFERLAALWRSGLTLDWSKLTGGEEGHRIPLPVYPFORSHLSSSLAAGHTPSSRPAVES
GAILAERSAGENAETRD CPLPTATLEPKAVAPAPLEATDAAGTRERLAE LWRELLGLTSIGPD
DHFFDLGGHSLTATRLRALIHQRFVDVLGLDEIFAHSRLSQLAARIEAAAKSRFSSIPSAPDQ
DDYPLSSAQQRIHSIVTRA EVGTAYNFFIVLELQ GALDRVRFEATFAALFRRHEGFRTFRVMR
DGGPRQRIVPDVAFRLPLTQVEPEQVPGRIEAFIRPFDLERAPLFR AELLQLAEQRHLLLFDM
HNL IADGISLNL FVADFAALYHGRPLAPLKLRYRDYAVWQEARLASDDLRSQREWWHRRLSPP
VATLALPPDFPRPAVRRYKGRNVVFHLDREIRDLVALARTQGV TMNMMLALWAALLHRETG
QSELVVGSLGGRPHSELHPVIGLFTNPLRLAVEGSTRFDRFLAACHQVFLEAYQRQDYPF
30 HLLVQELVPVRDPSRSPLFQTSLVYHNEIDGKTKLELEGLKVEVVPFEKGVARLDLKDVTPTF

5

139

10

SDRLECVLQYDLDFCEETMRGLIARFQALVAGLVADPAQSLAAASVSGKRALRAGVATASES
SPQSLPPQPSTAYATPSPQSPSPVVLTPADLPAILAAYVGQNPFPFAIHRGLILEAPLGLRA
LRSALDAVLGEHTHWRSVRAGDRARRVDKLELTSLVRLDDLRLVNPQANAFTLAWRDLAMPF
GEGRPLWRLRLAWSAPSRWLLLLTVHPLIGDNGTVDLFLAALADHLRRASAFPVAPLDEAELE
5 AELKWGEEGEGGLGLTAIAPVLGQLRESRLSPVAQMWLDEVCRRHDLTPLEVLAARLLDWTRSH
15 GHGSIALWTPLPEDHPLRDEGRCLQVRLLLEGPPSQRGAGDPSWLEQIALRRGTPATEVVCPTP
TQRAAIDLALAWLPQPPLHGLVGTVPWPESPLVCPFPLNLAFRPSHPPIAYALKHEATLAVTA
RARDLMRFLDGLGPES

20

10 pEPOcos6_ORF13.1 sequences:

(1) nucleotide sequence

25

Seq ID No 67 (>pEPOcos6_ORF13.1.seq)

ATGACGCAAGCCTCGGCCGCGTCGACGTCCCAGGTCGCGCCGGAGGTCACCCCGGCCGAAAG
15 GACGACGATGACGATCAAATCCGAGATGTCGGCCGTTGCTCACTCTGCGGAGAGCGGCTTCCG
CGCTGGGCCACGCGTGGGCGGCGCGATGAAGCGGGCCGGACGCCGGAGCAGGCCGCGTGAA
30 GCTGCTCCGCGCCCCGGTGAAGCGGAAGTGGCTGCCCCGGCGCCCGTCTGCGCCTGAGCGA
GCCGCGTATCCCGGAGGTGTGGGAGGCTACCGCGGAGCGCGGGATGACCCGAGCCCCGCCC
GCCGGCGCCACCATGACGCCGCCCCACGGGGCGACTCGTCCGGCGCGCCGGCGCGCCTCGGGG
20 CTTCCGCGCCCGGGCGGGCAGGTGCAGGATGGTCGGGCATGG

35

(2) peptide sequence

Seq ID No 68 (>pEPOcos6_ORF13.1.pep)

40

MTQASAASTSQVAPEVTPGRKDDDDQIRDVGRCSLCGERLPRWATRGRRDEAGPDAGAGRE
25 AAPRPGAEVAAPGARPAPEAAYPGGVGRLPRERGMTRAPPAGATMTPPHGASRPARRRASG
LPPPGGVQDGRAW

45

50

55

5

140

pEPOcos6_ORF14 sequences:

10

(1) nucleotide sequence

Seq ID No 69 (>pEPOcos6_ORF14.seq)

15

20

25

30

35

40

45

5 ATGGTGACGCGTCCGACGTCCGACGGCATCGAGGACGAGCTCGCGCCGTTCCTCCCGGTCCTG
CGCGGTGGCTCATCGAGGGCGAGCTCGGCCGCGCGGGATGGGGCGGGTGTTCGGGCGCGG
CACCCGAAGACGCGGGCGCGGGCGGCGATCAAGGTGCTGCTCGCGACTACGCCCGCGGCCG
GACGTGCTGGCCCGCTTCGGGCAGGAGGCGATCGCCGTCAACATCATCAACCACCCGGGAATC
GTCCGCGTCTTCGACTCCGGCGAGCTCGAGGACGGCTCGCCCTACATCGTGATGGAGTACCTG
10 GACGCCCGGGGGCTGCGCGACTGGGTGCAGGCCGTGCCGCCCGGGAGCGCCGCGGACGGTC
GTGCGGCTCGGCTACCAGATCGCCTCGGCCATGGCCGCGGCGCACGCGTCCAAGGTCTCCAC
CGCGATCTGAAGCCGAGAACATCATGGTGGTTCGAGGACGAGCTCGCGCCCGGGGCGAGCCGC
GTCAAGATCCTCGATTTCGGCATCGCAAGGTCTCTGGGGAGGTCTGCCCGAGGTGCTGGAG
25 CTCGAGGGGAGAGGCTCCCTCGCGCCCGCGTCCGCGTCCACGATCCGCACCGAGCTCTCGACG
CGGCCGCGCCGACGGTGGGCGCCACGACCGGCCAGAGAGCCCGCTGGGCGCGAGCGCCACG
CCAGAGAGCGCCCTGGGCGCGAGCGCCACGCCAGAGAGCGCCCTGGGCGCGAGCGCCACGCCA
30 GAGAGCGAGGCCCACGAGGAAGACGCGCTCCGGAGCCTCCCCGTCTGACAGCGGCAGGCCC
GCGATCCACCCCGCGCCGTCGAGATCCCGCCGAGGCGGTCTCTCCGCGGCGTCCGCGGG
TCGCGCGCGTCGATCGAGCCAGGCGCGCCCGCGCCGAGAGCGAGGGCGCGGACAGCCACG
20 ATCCCGTTCACGCAAGAGGGCGTGTGGGGCCTCGGGACGAGGAGCTACATGGCGCCGAGCAG
GAGCGCCACTCCGGGAGCGTGGACGTGAAGGCGGATGTCTACTCGCTCGGCGTCATCCTCTAT
GAGCTGCTCGAGGGGCGGACGCCCCGACGCGCCGAGCGCCGCGTGGCCGCCCCCGATGAGCGCC
GCCACGCCGCCCGATCTCGTCGCCCTCGTCCACCGGGTCTGGCGTTTCGATCCCGATGCGCGG
40 CCGCGCATGGCGGAGGTGGCGAGCGCGCTTCACCGGCTCGGCCGGGCGAAGAAGGAGCTCGAC
25 GAGGCGCTCTCGAGGTGGGTCTCGGCGGAGGGGCGCCGGGGCTCTTGCCGTGCGGCTATGCT
CTTCTCGAACTGGTCCTCCTGGGCCCTGGGAACCTATACGATTCTTCCAGCCTGTAAGTGCA
TTTCTCTTTCAATATCGTCCTCTCTTCATATACGAGGTGAGTTCTCTGAGGTCTCTCTATAAG
45 TCTGGGGTGTCTATTTCGGCCTCTTACTTGTTACTTCGCCTCTTAGGAGTTTTCTCTAATT
TTGCCCTCTTACATTCCCGTATTCACTTAAGTGGGCCCTATCTCATTCGC

30

50

55

(2) peptide sequence

Seq ID No 70 (>pEP0cos6_ORF14.pep)

MVTRPTSDGIEDELAPFPVLRGWLIEGELGRGGMGRVFRARHPKTRARAAIKVLLGDYARRP
DVVARFRQEAIAVNIINHPGIVRVFDSGELEDGSPYIVMEYLDGRGLRDWVQAVPPAERPRQV
VRLGYQIASAMAAAHASKVVHRDLKPENIMVVEDELAPGGSRVKILDFGIKVLWGGLPEVLE
LEGRGSLAPASASTIRTELSTRPAPTGVGATTGPESPLGASATPESALGASATPESALGASATP
ESEAHEEDALRSLPVVTSGRPAIH PAPVEIPPEAVSSAASRGSRASIEPGAPAPQSEGAGQPT
MPFTQEGVWGLGTRSYMAPEQERHSGSVDVKADVSLGVILYELLEGRTPDAPSAAWPPMSA
ATPDLVALVHRVLAFDPDARPRMAEVASALHRLGRAKKELDEALSRWVVGGA PGLLP CGYA
LLELVLLGPGNLYDSFQPVSAFFQYRPLFIYEVSSLRSSYKSGVSYAS YLLLRLRSFSLI
LPSYIPVFILTGPYLIR,

or DNA sequences complementary to said open reading frames,

(b) DNA-sequences which hybridise under stringent conditions to regions of DNA sequences according to (a) encoding proteins or to fragments of said DNA sequences,

(c) DNA-sequences which hybridise to the DNA-sequences according to (a) and (b) because of a degeneration of the genetic code,

(d) allele variations and mutants resulting by substitution, insertion or deletion of nucleotides or inversion of nucleotide segments of DNA-sequences according to (a) to (c), wherein the variations and mutants offer isofunctional expression products.

10. DNA sequence according to any of claims 1 to 5, wherein the DNA is selected from the group consisting of

5

142

(a) the following DNA Sequence:

10

Seq ID No 71 (>Contig43)

15

20

25

30

35

40

CGGGTATTTGTGATATGTGGGCNGTAGTCGTATGCTTCATTAAGTACATC
5 CGTCCGTNGTAGAGAGTGA CTCTGTGCGCAGCGATAATAGACACGCTTGTG
ATGCTATAGGGAACATAGAGTCNTAGTAGATGATACGACGAGATATTNGT
ATAGAGCGTATAGACCGACGTGTGAGCGTCATAAGTGTTGTGTGTCATGA
GTGTGCTCAGAGGACGTGCAGACATTATATGAGCAGATCATGAGAGAGAA
TCAATGCTGCAAGNTATTCGTGCAATCTACATTATATCGAATCGTGTATG
10 TGCCTTTGTGCGCAGCGGATNCGATGAGATACCGAAAGGGTATGTATCTA
TNTTCGTGACGCTCGATNAGAGCAAATCCGCTACCGTGGAGATATCGTGT
ATCGACTCCATCAGCATCAGTATCATGATACGTCAAACGAGTACACTCAT
TATTGATAACACACGTANGTGTGCATGCACAGTTATCGAGTGTATTGTGT
GCATGAGAGGTATAGGATNTATAGGCGAGCATATATATCTATATATATAG
15 GTTAAGAGTAGAANACTATGAAGATGCAGGAAGTAGTATCTCGCGGACAA
ACGNGTAGCTAGCGGGGTGAAGTATTATCGACAGTGTATAACGACTCA
ACAGGNTACGAGGTACATTGTATTTACAGTGTTGGAAGGATTGCGCGA
GGAAAGGTAGTGGTACCGTGTGAGCTACGATGCTCGGGATAATGGTGATT
AGATAGAACCCTTAGCGTTGCTAGATGAGTGAGTGGTGGTATGAGTAGAGT
20 TTTTGTCTAGCTTTGTGTCCAGCGAGGATTCGTTTCAGTCTGAAGGGTAA
GAGTACGTCCATCGCACACCCGACCGTTTTGAGGAGTTCTCGGTGCGTGG
TCAGTGGGGTTTGGAGAAGACAGAGTTGATTCATAGGGTTATCAAACGAG
TTATGTGGATAGATGGTAGTGACCCCATTTGAGTGAGAGTGTGGCGTTA
ACANCAGCAGGATNTAT

25

SEQ ID No 72 (>Contig44)

45

30

TAGTCTTTGACACCATGGGAGCTGCTACCGATGTTGCCGAGCACGATCG
CGCTGGCGCCGACGAGCGACTGCAAGCCGGCCGCGGCCATTTACGCCTGA
CGAGCGAGGTGGGCGAAGTGCTGGTGCGCGCCGTGCGTGTGAGCGCGCC
CAGTCCCGCGTTGCGCCGTGCGCCGAGCAGTAGCGCGCCGTGCAAGACGA

50

55

5

143

10

15

20

25

30

35

40

45

50

55

TCACCGCGATCGAGGTCAGCGTCGTGGGGGCGAGGCCGAGGAGCGCGAGG
ATGCCGAGCACGACGCCGGCCGCGCGTAGACCAGCTTGGCGCCCATGCC
GCCGCCGAGCTC3GTGCGCGTGTCCCACTCGACGGGCGGCGCGGTGCGGC
TCAACGCGCCGAAGCGCGAGGCGATCGCGCCGCCCTGCGCGATCAGCGCG
5 GCGCCGAATACGATCGTGGCGATCTGAGGTGAGCTCTACTGGCATGATCC
15 CCGTCAGCCCCGAGGATGGTGAGGACAATCGTCGCGGCGCCGCACAGTACC
TCGCACGAGCGAGCCTCCGAGCACGACCTTCGGCGTCGTCTCGTCCTTTG
GTCTGCGTCGCGCGCCCGAGTGCGGCGTTATGTGGCTCTCCGGCTGTGCA
AACCGTTCACGTTCTTCCGGTCTTGAGTCAGCATCGGCATGATTCCCCC
20 GTCTGCGGTGAGGCCCTTGTGCGGCTCACGCGGCTCCGACTTGACGCTG
10 CTGTGCCGGGTCTCTCGCTCAGGAGGCGCCTCTCTTGGTGGTGCTTGCG
TCCTGGTCCGTTTGCCCGCTGTGCGGTAGGTTTCTTGAACCAGGTGACC
TTCAGGAGCCCTTGATGCGCTCCATCGTGTCTATGTGATCCTTCTCT
25 GACTTGTATGGGTCTCGAACCAACTACGCTTGATCAGGCCTTCGAAGGGT
15 CCTTTGGGAGATCGACTCTGGATCCATACCGGGAGCCCTGTTCTGCCGC
TCTCTTAAGTTTCCCTTCTGTATCCGTGTGACCGGAAACGCTTTATCT
30 CTAATGCGCTCTAATTGCGTCTCTGCCACACGTGCGCTTCACTCTGGATC
TACTTCTTCTCCCTAGTCTTCTACCTCCGTACCCTTATTGTTGGTTCTA
TTTATTTCTTTTCGCTTCACCTCGCGTCATTGTGCGCTAGTGTTCCCTCCC
20 TCATATCGCCTTTGGTCTCCCTCGAGCGTACAGTCCTCTCTTTCAGATG
35 CTTTCCGGCTCCTCTTCTGCTGGCCCCCTTATCCTTTCTAATACTTC

SEQ ID No 73 (>Contig48)

ATGCGCCAGGAACACCCCGGTGCGGCTGCCGTGAGGGACTGGGGTGCG
40 ATGCGGCGCTCCTCGAGCCCTTCCCAGGTGACCTCCAGCAGCAGGCGTTG
25 CTGAGGATCGAGCGACCGCGCCTCCCGAGGCGAGGTGCCAAAGAACGCGG
CGTCGAAGCCGTCACCGCCTCGGTGAGCAGTCCGGCCCAGCGCGGCACC
45 TCCTCGCTGGGATGGACGCCGACCAGCGCCAGCGCCGCTCGAGCGGCTG
GACCGCGTCTCGGCCTGAGTCGAGCAGCTCCCAGAATGCCTCCGAGTGT
30 CCGCTCCGCCGGGAAGCGGCAGCCAATGCCTACGATGGCGATCGGCTCG

5

10

15

20

25

GTCCGCTCTTGCTCCAAAGACGCGTTCTTTTCGCAAGCTTGTCCATGAG
CAGAAGGGCATGCTCAAGCTTCCCGGCATTTCGTGGTCGCCATACTCCCTC
GGTCCCTTACTCACCAACGATCTGCGCGAGCTGCGCCAGCTTTTCGGCGA
GCAACGCGTCCTTCTGCTCGTCCGTCATGCCCCGAGAGCCTCGAGATCT
5 GCGGCATCGTTCTCGAAGCTCTTCTCCCGCTCGGTGGCCGAGCGTGGGT
10 CGCGCCGGCATTGGAACAGAAATGTCTAGCAAGCTCCCGCTCAGAGCTG
CTACGTTAGGGTAGGTCCATAGCAGGGTCGCCGGCACGGTGATGCCGAGC
GCGGCCTCGATGCGGTTGCGGAGCTCCAGGCCTATCAGCGAGTCCATGCC
GAGATTGCTGAACGGCACGTGCCGCTCGATCCTCTCCGGCGGAAGGCGCA
20 GCGCCCGCCCCAACAGCTCGCTCAAGTGCTTCTCCAGAATCAACTGACGA
10 TCTTCGGGCTTGGCGCTCTGCAGCGCCTCGCGCAGGTTGACGCGTTCGA
CGCGCCTCGGTGCGGCGCGTACGCTCCTTCAGCAGCTCCGCCCACAGCG
CCAATCGGGCCGCGTTGGGATAGAACTC

15 SEQ ID No 74 (>Contig49)

30

35

40

45

50

ACCACCGCTTCACTCAGTATGTACTTTGTTATACTCGTCTTAGTACAATG
ATATAATACTCATGTGTATTCTTAATCTCGGGGAGANAAAATTGGAATAC
TGGACACCGTTGCCGCATGCNCACTCTAGAGATCCCCCTGCGACGGTATC
CCACGGCACCGGTATGGCCGGCGCGCTCCGGGGGTCAACGCCCCGTGG
20 TTGCCTTCACGACAACCGCGTCCGGCGGGCGCCGTTTCGATGCCGCGGG
CCC GCGCGGGCGGCGGCTTATCCTGTGGAGCATCTGGAGGGCGCTCACG
CACCTGTCACTCTAGTTCTGGCCCGCCCGAAGGAGTCCGGGAGGCCGAA
GTTGAACCCGATGTAGAGCGCGATGAACGACGGGAGCACGCGCGCGGGGA
40 TGTGCAGCGCGGCGCCGATCGGCGTCGCGAACAGGACGAGCTCGCCCGGC
25 ATGCCGGGCACGACATACCCGAGCAGAAACAGATCGGCACCACGAGCGT
GAGCTCGAGCAGCGATATTTTCATGACCGACCGCGGGCGGCGGCGCCCG
CCATGACGAACACGAGATCAACGTGCCGTTGACGTTGAGCCAACCGCCG
AGTCCCACCACGAAGAGCCTGAGCTCCTGCGGCACCGCCGGATAACATTT
45 GCGGACGAGGTGCAGGTTGAGCGGCGTCGCCAGCGCCTCGCTGCACGAGG
30 CCCACAGCAGCGGATAGACCTTGAGCCAGTAGTTGACGAAATAGTCGCGC

55

5

145

10

15

20

25

AGCGAGAACTCCGGGGCGGCGGCCTTCATCCGCAGCAGGCTCGCCGCATG
GAAGACGAGGCAGGCAGCGCCGACGACCCCGGACACGAGCAGATAGGAGA
GCATGAGGTCTCCGGCGCCGAGCGAGGGCGCCCCCGCCGAGGCGCGCGCG
AGGTCCCCCCGTGCACCTGCGCGGCGAGCTGCGGGGAGCCCGCGGAG
5 ATAGGCGCCGAGCCGAACATGAAGAGCGGGACCAGGCACTGCACGGCGC
CTCCCGCGCGCTCCAGCGCGTCCGCGCGCGCTCCAGCGCGGGGCGACC
CGCGGCGCGCGTACGGCCGCGAACGACGTACGATGCCGCGGTAAAGGGC
GAGGAAGCACGGGCTCGAGATGACCAGGCCCCGAGGCGCTATAGAGGTGC
GCGCGGCTCGAACGGCGCGCCCGTGTGTGGCTGGGGAGCAGCGGCAGC
10 CCGAACACGAGCCATGTGACGACGACCCCGAACAGGCACGCTGCCAGACG
CTTGAGGGCGAGCCAGCCCATGATGACGCGAGCAGCCCGCCCGGCGCC
CTTGCCGGTGCAGGCTCACGAAGGTGGCACGAGGACGACGAAGATGACG
ACCGGCGCCAGCGTGGTGTACCAATGCAGGAGACCGTCCATGGCGCGGGT
25 CGACCACGCGTGACGCTGGTCTCTGTCTGACTCGATCATGGCCCAT
CGCCTAAAACTAATGATCCGTTCTCAAATTGGTCAAAAAAAGTTCCCTT
AAGACTGTTTTACTCCGGAATATTAATATATTTCTGAGTGTGAGGTGATG
30 TTAATCACACATTCTGATATTCTCAAGGGGAATCCGTGTCATTGTGAATA
CTTCTCTCTCTACAAGAGAGGTTATATATGGTCTCGAATATCTCGTCCGC
TCTTATATATATTCTCTGTGATAATATATATCGAGTGTGGGTACTCAGC
20 TCTCTTGGTGTAACTATAACTCGGCATCTCTCATAATACCTTATATATA
CACACTCTCTCGGTCAATCTCGCATAATAGATATATTTATATGTTCCG
35 CGTTTTATCCGAGTGGGATACACTTTTCTATATTTCTTTGGTGTGACG
CGTGGCGTCGAGCCTTATTATTGATTGGTAGTCACGATATTCTCTAGAT
40 GACATCATACAGATGCTCATAACTCGATAAACACAGGTGCTACACGACGA
25 GACTCTCACTCTCACTCTT

45

SEQ ID No 75 (>Contig50
TCCCCAGTTTTCTCCTCTCTACGCTCAGCAGGAAAAANATAAT
GGAGAAATCGTTGCGCTCTAGCAGCACTATAGGATCCCCGCTGCTCTTCT
30 TCATGCACCTCGTGGAGCAGAAGTTCATCAACGCCTTCGCGATCATCGTG

50

55

5

10

15

20

25

30

35

40

45

50

55

GCGGTGAGCTTCCTGGCGTTGCTCCTGTGCTCGTCTCGTCCGACGTCCG
GACGCGGAACACGTTCCCGCCCGCGCCTTTGCCGGCGCTGAGCCCGCCGG
CGCCGGCCTGATCCCGCCGGTGCCGGGCGGATCCGTTGGGCCGTGCGCG
GAGCCGCTGTGCTCGGCCGGTGATCGGTTGTGCGGGCGCCGTGCCTCGG
5 GCTTACTACCCCTCTCGCGGGTGGGGATATGGCCGTGGATGAGGGAGGC
15 GATGAAAATCGTGATCGCCACGTGCGCGTTGTTCTAGATCGTCCCAGGCT
GACCGTCGGGAGCGCCAGCACGAGATGAAGAGCCACACCGCGAGGACCG
TGTTGAGGTACCGCACCGCAGGGGCGAGCATGGCGGTGATCGCGAAGATC
ATGCAGAGCAGCCCCAGCACCCATGTGTTGCTCCGCTGCGCGTGGCTGTG
20 CGGCCAGATGACGGCCGAGATGAGGAGCCAGAACCCGAGGACGACGTTCA
10 CGATSCGCGCCATGAGATTGCCAGCTCGAACCATGCTCCCTCCACCTCC
GATCATGGGACCGATCGGGTCGCCACGGATCGATAACGGGCGTCAGGAGA
CCGTCAATCGGCGAGCTCGTGAGCCATGGCGACAGCCCGCCGACCGCGCC
25 GCGGGGTCTCTGGCCTGCTGGTCGCCGTGCCGGCGGCGCGATGGGCCTG
15 CCTCCGGTCGGGCCGCGCGGGCGCGGGCGCTCCGCGGAAGCTGGAGAC
GCGGGACATCGTCCCTCGCCCCGCGCTCGGACGAGCGGCGCGAGCCACT
30 TCTCGACGGCCGAGCGGGCACTAAGCTTCCGTCATGAGGCTCGGCGCACG
GCTCACCACGCACACGTTCTCGGCCGGCGCCGCCGCGCATCAGCTTCGTG
TCCAGCCGATCCCGGGCTCGGACCAGCTGTTGTCATTCCGATCCAGTAC
20 CTGCTCGGGCGTCGCTCGCGAAGGAGCGAGGCGCGCCGCTCTCGAAGGC
35 GCGGTGGTCCCAGGTCCACCAGCTCATCTGGGGCGGCGGCGCTTCGCC
TCATGCTCGGCTTGACCCTAGGGCTGATCCCGCTGGCCGGCGCGTTCACG
AACGCGATGACGGCGTTCTCAGACCGAATATCTCGGGTACTACGTGGA
40 TAGAGCCCTCGACAACCCGGACAATCCGCCTCCGGCCCTGTGATCCAGG
25 ATGTCTTGAGCGCCATCACCTCCTTCTTACCGGGCGAGCGCGGTAGGCG
AGC3GTCCCTGGGTGAGCCCCACCCTGCGGCTCTAGGAGCCGAAGGGCGA
GCTCCTCGGGAGCGGCGCGCGTCAACCAGATTGCGCGGCGCTTGCGG
45 CCG3AGCGTATCGCGACCGCCGCCACCGCCGCCACGGCGAGCACGGTGAC
CG33GCGGCCCGCGCATGGCGACGCTCCGGGCCGTGTGCTCGGCCTCGC
30 CCTTGATGCCGCCCCACCTCGCCTTGACCGTGACCAGGTGCGCGCGGAGC

5

147

10

15

20

25

30

35

40

CGCTCCTTGCTGTGCTCGATCTGCTCGGCAGGCCCGCGGGCGCCACGAA
ACCTGCGCCGGCGCTGACGTGACCTCGCTCGCTGAGGGGACTGGCGACCC
TCTCCGTGTGGAACCGGATTGCGCTGGATGGATCCATATGTCCCGCGCTG
CAATCGTTGCTCCCCGCCGGCACATCGGAGTGCTCGCCGGATCGCGCGGC
AGCGCCGACGCCGTACTTCCATAGGATAGCCCACCCCATCGGACAAGCCG
GCTCCTGACGGCGGGCACCGAATGTTCCGCCAGACGGGCACAAGGCGCACG
CCGCGGACGGATCGGCCGCACTGGCACTCCAGAGCGCATCGACGGATGGC
CGACGGATGTGCAATGAGGCGCCCGCACGAAGCGAATTGTCCCGAATACA
GCGAAGAAATCTATAGCGATGCGAGCAGAAGGATATGTCTATGGGGGGCA
GTGCAAACTGGGGACAGTCAACGCACATATTCTCTCCAANTGCTAACGA
CAGCGTGCGCAGAGGAAGTATCCTACTAGTGTAAGAGGGACATTTCGATGC
GACCGCATAAACATTTCAGTCTACAACCGGTGAGAGGATGGAACACCCCGC
CCCTCTGAAGGCTAGACAACCATGAATATGTGCAGAGGAAACACAGAATT
CCAAAGGTGAGAACATATGTAGGATCGCGCCACCCGAGATTGAGTGAAGA
TATACATATATACTTATATGGATCTACAACATGGCGAACCGAACGTAGCA
NAATAGTAGATATAATTGTAATACTGAGCTACCGACAGAAAGATACACAC
GAGTGTAACACATCACACGCAGAGTGGTACCAAATTCACACCATGCGAG
CCACAATGTGACACGGAGGAGCACAGCATGGGCGCCACTATGGAGGAGAA
ACTACTGCAACCCACATCTGATGGACTGACCGCACGGACGGGACGTGTCT
ATACATACAGATACATCNGATGGAGGAAGATGCATGTGCGATGATATCAT
CGTCGCAAACTCATATGTGGAAGAAGATATGNGTCAACTCAGCACTACTC
ACACGATACGTGAACAGGAGTGACTAGGACATCNCATGGTGTGTCGGCGC
GTGCACGTGATATCAAACTCTCTGATCAACCACACACTATATAAGGAGTA
TCGAGCGGCGATGGAACACCCCTCACAGCATACGTATATGCACAACGTC
TGAACACTCTNGAGACACAGTGAAGG

45

50

55

SEQ ID No 76 (>Contig51)
GATCCAGTTACGCCCCGCCCTCGGTACGCCGGGGTTTTTCGGCGTCGA
CCGCGGACGGTCGGGGCGACAACCGGGGGTTGTGCTCAGCGGTTTCGGTG
GATGTGCGCGACGAAGTTGTCCGCCGCCGAGTCGTCTGCGTGCCGCCGC

5

10

15

20

25

30

35

40

45

50

55

GCTGCCGTACGAGCAGCCCGAGGCACCGCGCGAGCTCCTGCCGCTTCCGA
TGGCGGGCGGCTGCTCGGCCTCGTGCTCCGCGATTCCGGCGCTGGCGCCT
GCTCTCCTCGGCGCGTTCTCGCTCCGCGCGAACAGCACCTCCTGCACCT
CCGCGAGTTCTCGGCGGTAGCGACCCGCACGTGATCCGCACGTCTCTCCG
5 AGACCGGTGACGACGTCAACCGCGGAGTAGTACGCCTCCGCGCGGACGAC
GTCGTCGAGGTGAGATTAGCGTCCGCTCCGACGACCAGGCGCATCGCGT
15 CGGAGGTGAGGACGACGCCACCGTCGAGCGCGCGGCCCTCCGGCTCGTAG
ACGTA CTGGAGGTCCGTCTCCTCGTCTCGTCACCGTCGAGGTTCGACGTG
CCGCCGAGGCCTCGGGTCCACTCGATCGCGCGTCTCGCGCGAGGGCTT
20 CCTCGTACTGGGCCCACCACGCGCGCAGCTGCTTCGGCGTGCCGTAGCCC
10 TCGGCCATGTGGGGTTCGAGCCCGCGACCTCGATGTCCACAGTCGGTA
GAGGATCTGGAACGGCGTCATGGA CTTCGGGCCCCGGCCGTCTTGGAGT
CCAGACGGCGGTCTCCATCGCAGCTGCGCCGGCGGCTTCGAGGTCTTGG
25 TCGACGGAGTCGGGCGCTCTCGCTTCCCGTCTTGGTTCTTGGTGAGGTA
15 CTGATCAGCGCAGTCGTCAGCTGACCGGACGATCGAGACCATCACGC
CGTGGCCCTTGCCCTTGCACTTGACGCCGGGGGTGTCGAGTCGGTCGAG
30 GGCTCGAACTTGGGGTCAGCCCGCTTGAGGGCGCCCGCCACATCTCCCG
GAGCCAGTCTCCAGTCCCCAGGTCCGTCTCGAGGGCTCAAAGTGTC
CGACGACGTACCCCTTGGCCGGGTGCCAGAGAGCTCGCCGCCGAGGAAG
20 ACCAGCAGGTTGAGGTGGGGTGGTAACCGTTCTTCTTGGACCGGTGAC
35 CTCAGCCGCGCGGACCATGCCGATGTAGCCGATCCGGTGGCGGATGCCGT
CCTCAGCGGGACGGACGTACTGCGTTCCGTCTTCCGGGTGCGGCGGGCC
TCAGGGCGGCCGTAGAAGGCCGGGGCCGTGAGCATCCGCTGGTAGGCACC
40 GGGCGCGCGGCGGGCTTGCCCGACCGGTGAGGACCGGGGCGCCCTTGT
25 CGTCCAGGAGAGGCCCGCCCCAGAGCGGGCGACCGGTGTCGAGGTGCG
GTGGTCTGGTTATGCCGGGCGGTGAGGACGACAACGGCGAGCGTGCCGCC
GGCGCGAGGTGCCGAGAGCACCGGTCTTGATCTCCTCGGTCCGGCCAC
45 GGCGGATCGCGGAGGAGCACTCCGGGCATAACCAGATCCGCCCGCAGCGG
ACCAGGCCGATCGTGACGACGTACCCACGGCTCGACTTCGCGTAGATCAC
30 GCCGGTGTCCGGTTCGAGGACCGTCGCCCCGACCCGCCGAGGCGTCGA

5

10

15

20

25

15

30

35

40

45

50

55

TCCCGGAGACCCGGTTGAGCACCTTGCGGCCCTGGTAGCGCGTACGGCA
GCGGTGCGCCGCGCGCCTTGTGGTCTGTTCCGAAAGGGCTGCCGCCCTCTC
GGACTCTCCCGTTCTTCCACGACTGCCACTTCCGCAAAGTCGCTGGTCA
GTGGGGGGTGGGAAACTCTGTCAACCCTTTACCTAGGCGTCCCTTTTTG
CCAGGGGCGGTCTCACGGGCGGCCTCGGCGGCTCGGTGCGGGCCTTCCG
GGCCCCGCGGGCCCGCTTCTTGACGCCTCGGAGCAGAACTCCTTGGCTC
GCTTGCCGGGGGTGATCGTGAGGGCGGCGCCGACGCGCAACGGGGGCCG
GCGGGACGCGGGCGGGCGACTGAGTCGGCGCGCCGATCAAAGAGGGGGT
TGCGGACGCCAAAGCGTCCCTTACGCTGGACACAGACGAGTACCTTGGTT
GGTAGCCGGGTGGACGTGAGAAGCGGTGAGGATTAGGACCCCTGGCCGT
TTCGCTTTTTCTGGAGTTGTTGCGGTAGATCCTGCCGCATCGCCGCCTC
ACGCGTGGCTCGCCGCGCGGATGCCTCAGAGGCCCCACCGGTCGTGAGGA
CGCAGACGTGCGCGTGTCTCCTGGTGGTGAATCACCAGCTCGACCACACGG
GCGCGGCCGGCGACGTCTC

15
SEQ ID No 77 (>Contig52)

CGGGATCTGGCCTTCATTAACCAACGACGGGGCAAACATAATAGGCTGGG
CATTGCGCTTCAGCTCACCACAGCCCGTTTTCTGGGAACATTTCTGACGG
ATTTAACTCAGGTTCTGCCTGGTGTCAACATTTGTGCGGTACAGCTT
AATATCCACCGTCCAGAAGTTCTCTCCCGCTATGCTGAACGGGACACTAC
CCTTAGAGAACATACTGCATTAATTAAGGAATATTACGGCTATCATGAAT
TTGGTGATTTTCCATGGTCTTTCCGCTGAAGCGTCTGCTATATACCCGG
GCGTGGCTCAGTAATGAGCGACCGGTCTGATGTTTGATTTGCCACTGC
ATGGTTGCTTCAAAATAAGGTATTACTGCCCGGAGCAACCACACTAGTAC
GTCTCATCAGTGAATTCGTGAAAGGGCAAATCAGCGGCTGTGGAAAAAG
CTGGCCGCACTGCCGAACAAATGGCAGGCAGCTCAAGTGATGGAGCTTCT
GGTCATTCCGGAAGGTGAGCGTGTATCAGCACTGGAACAGTXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXGCTGGAACGATAT
ATCCGATTACGAAGTCTTGAGTTTCCCGACTGAACTTTCCGGTCTGCC
TGCCATTCAACTGCGTAATCTGGTCTGTTATGCTGCCATGGCGTCGGTAA

5

150

10

5

15

20

10

25

15

30

35

20

40

25

AATATATCGCTCGAATGCCACAGCAGAGAAAGCTTGCTGTACTTACTGCA
TTCGTTAAAGCACAGGAAATAACGGCATTAGACGATGCCGTTGATGTGCT
TGATATGCTAATTCTGGACATTATCCGCGAAGCAAAGAAAACCGGGCAAA
AAAAAAGACTCAGGACACTGAAAGATCTTGATCAGGCCGCATTGTTACTG
GCGCGGGCATGTGCATTGTTGCTGGATGATAATACAGATGTCCAGATCT
CAGGCAGGTTATCTTCAAGTGCCTACCCAAAAACAGACTGGCAGAATCTG
TAAGCAAGGTTAATGAACTTGCTCGTCCACAGAACAXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXAAACGTTTTCTCCGGCGGTGTT
GCGGGACCTGCATTTCCGTGCGGCACCGGCAGGTGAACATGTAAGGCTG
CGATTCATTATCTGGCAGAACTGAATGGTTCGAAAAAGCGCATCCTTGAT
GATGCGCCTGAACATATTATCACCGGTCCCTGAAACGCCTCGTATACGA
TGCGGAGGGACGGATACAGCGTGCAGGTATTCACTATGTTTGTGGAAC
GCCTTCAGGATGCACTGCGCCGCCGGGACATCTGGCTTGAAAACAGTGAT
CGCTGGGGAGATCCTCGCGAGAAGTTGTTGCAAGGTGAAGAGTGCGAGAC
TCAGCGTATTCTGTCTGTCTGGGCACTGGGACATCCTGTGATGGACGTA
AAGGTGTGCAACAACCTGGCTATTTCAGCTGGATGAGACCTGGAAAGCCGTG
GCATCAGATTTGAAAAGAATGCGGAAGTTCATATCTGTAATGAAGGTAA
ATATCCATCCCTGACTATCAGTTGTCTGGAGAAACAGGAAGAGCCACCAT
CATTGCTTCGTCTAAATAATCGGATCAAACAGCTACTCCACCGGTAGAT
TTAACGGAACCTGTTACTTGAGATAGATGCCCAGACAGGATTTACACATGA
GTTTGCGCATGTGAGAGAATCTGGTGCTCGAGCGCAAGATTTGCACATCA
GTTTATGTGCGGTATGAATGGCTAAGCCCTGTAATATGGGCCTGAACCCG
TTGATAAAGCACAAATATACCAGCATTGACCCGCCATCGGCTCAGTTGGGT
GAAACAGAATTACCTTCGTGCAGAAACGCTGGT

SEQ ID No 78 (>Contig53)

45

30

ATTCCACGCGCTCACGGTCAGCTTCGACCCGCGCGAGCGCCCGGCGCCG
CCTCGCAGAAGCGCGCGGTACGCTGTCCGAGCTCGGCGCGGACGCGCAG
GCGCCGGAGTGGCCGTTCTCGTCCGCGACGAGGCGGCGACCCGCGCGCT
CGCCGAGGACCTCGGGTTCCGCTACGCCTACGATCCGACCACCGATCAGT

50

55

5

10

15

20

25

30

35

40

45

50

55

ACGCCCACCCGGCGGCCGTCTTCGTCTGACGCCGGACGGGCGGATCTCC
CGGTACCTGTACGGGACGGAGTTCCCGGCGCGGATCTCCGGCTCGCGCT
CCTGGAGGCGAGCCGCGGCCGTATCGGCACGATCGTCGATCGGGTGATCA
TGACCTGCTATCGCTTCGACCCGGCGAGCCGGAGATACGCTCCGTTCCCTA
5 CTCGGCTTCCTCCGGCTCGGGGCGGCGCCATCCTGATCACGGTCGGCGG
GCTGCTCGCCGTCTGTGGCGGCGGAGCGCCGGCGGCCAGGTGCTCGCA
15 CGAGCGCCGCGTCGGTCGTGACCCGTGGCCGACCGCCAGGGAGGTCA
CCATGATCAACGAGCTCCTGCGCAAGCTTCTTTTCTGTCCGGCCAGTGG
TCGACGATCGTGTTGACATTTACAAGCTGCTTTACTTCGTGATCTCGGT
20 GACGATGGCCGGCGGACGCTCGTCGCCCTGTTCCGGCCTACCTGATGA
10 TCCGGTACCGCAGGCGCCAGCGGGATGTTGAAGGCCCGTTCCCGGAGCG
ACCGCGAGGCCTCCGCTCCTCCTCGAGGTCGGCATGGTGCTGGGCCTCAT
CGTCCTGTTCTCTGTCGTGGTGCATTGGAATGCGGCAGTATGCAGAGC
25 TCCGCGTCGCCCCCGCGGACCCGGTCGTGGTGTACGTGACCGGGAAGCAG
TGGATGTGGAAGTTCGCCTACCCGAGGGCCCGAGCTCGGTGGCGACGCT
15 CTATGTGCCGGCGCGTCGGCCGGTGAAGCTCGTCATGACGTCCCGGGACG
TGATCCACAGCTTCTTCGTCCCGATTTTCGCATCAAGTACGATGTCGTC
30 CCCGGCCGCTACACCACGCTGTGGTTCGAGGCGACCGCGCCGGCGCCTA
TCAGATCCTGTGCACCGAGTACTGCGGGACGAACCACTCCACCATGCGCG
20 GCGAGGTGATCGCGCTCGAGCCCTCCGATTTCCGCGGGTGGCTCTCCGAC
35 CGCGGGCGGGGCGCCGGTATCGCCGGACAGGAGTACACGCCGCCGTCGAC
GCCGGGCGAGGGGATCCCGCGGAGCCGCTCAGCCTCGTCCGGCTGGGCG
AGAACATCGCGGCCGAGGAGGGCTGCCTGCGCTGCCACACGCCGGACGGG
40 ACACCGCACATCGGGCCGACCTGGGCCGGCCTCTACATGTCGGTCGTCCC
25 GCTGGAGAGCGGCGGCGCCGGTCGCCGACGACGCGTACATCACCGAGT
CGATGATGGATCCGCTCGCCCCGATCCACCGCGGCTACCAGCGGGTCATG
CCCTCGTTCCTCGGCCGGCTCCAGCCGGCGCAGGTGCGCCCATCGTCGA
45 GTACATCCGGTCGTTGAGGGGCGTCGCGCCGAGCCGGGCGCGCGGACGC
CGCTGCCCCAGGGGCCCGCCCTTCTGCGCTCCGGCCCCGAGCGCCCCGCC
30 CCGCTCAGCGGGGCGCGCCGGTCGGCCCGATCGAGGGCGGCAAGCCCCG

5

152

10

GGAGGAGCTCCGATGAGCACGGAAACGTACGAATCTCTGCCCCGACGCGCC
GGCCGAGAGGCCCGAGCCCCGACTACCTCCATGTTTACCGCGGGGTGACG
GAGTGGCTCAGACCACGGATCACAAGCGGATAGGTCTCATGTTCTACGC
CGTCATCGTCGGGAAAGCTTCTCTCTCGGAGGCATATTCGCCCTCATCAT
5 GCGGACCGAGCTCCTCACGCCCCGAGCGGACCATCATCGACGCGGCGACCT
15 ACAACCGGATGTTTACGCTGCACGCGGTGATCATGGTCTGGCTGTTTATG
ATCCCGTCGATCCCCAACGCGTTCGGCAACTTCGTCTCGCGATCATGCT
CGGCGCCAAGGACCTCGCGTTCGCCCGGATCAACCTCGCGAGCTTCTACA
TCTACCTCCTCGGGGCGGCGATCGCGATGGGCGGCATGATCGCGGCGGC
20 10 ACGGACACCGGCTGGACGTTCTACCGACGTACAGCCTGAAGACGCCGAT
GACGCTGTTCCCGGTCGTCTTCGGCATCTTCATCGTCGGCGTCTCGTCCA
TCATGACGCGGGTCAACTTCATCGTGACCACGCACACGATGCGCGCCGAG
25 GGGCTCACGTGGAGCCGCTGCCGCTCTTCGTCTGGAGACCTACGCGAC
GAGCATCATCCTGCTCTTCGCGACCCCGGTCTCGGGCTCTCGATCCTGC
15 TCATCGGCATCGACCACGTGACCGCGCTCGGGATGTTTCGATCCCCGGTTC
GGCGGCGATCCGGTCTCTTCCAGCACCTCTTCTGGTTCTACTCCCACCC
30 CGCCGTCTACATCATGATCCTGCCGCGGTTCGGCGTGGTGAGCGAGGTCG
TCTGACGTTTCGCGCACAAGCGCCC CGCGTCTACTGGGCGATCGCCATC
TCGTGCTCGGGATCGCGTTCGTGGGTTCTGGACGTGGGGCCACCACAT
20 GTTCGTGGCGGGGATGAGCGAGTACGCCGCGGACGTCTTCGGCGTGCTCT
35 CGATGTTTCGTGGCCATCTTCTCGGCCATCAAGGTCTACACGTGGGTGCGG
ACGCTGTACAGGGGCTCGATCCACTTCAACACGCCGCTGCTCTACTTCAT
CGCCTTCCTCTTCTGTTCTGTTCTTCGGGGGATGACGGGCGTGGCCGTCG
40 CCACGCAGTCGCTGGACGTGCACTGGCACGACACATACTTCGTTGTGGCG
25 CACTTCCACTTCATCATGGTGGGCGGGACGCTCACCATGTTCTCGCGGC
GGCGCACTACTGGTTTCCGAAGATGTTTCGGGCGCCTCTACTCGGAGCGCG
TCGGGCTCCTCTCGGCCGCGTCGGTGTCTCTCGGCTCTTCTTGACCTTC
45 TTCCCGCAGTTCTCTCGGGAACATGGGGATGCCCCGCGCTATTACAG
CTACCCGCGCGCTACCACTGGCTCCACGTGCTCTCGACGGCGGCGCCT
30 ACCTGCTCGCCGCGGCGCTCGTGAATCTCGCTCTGAACCTCGTCATCGCG

50

55

5

10

15

20

25

30

35

40

45

50

55

CTCAAGTGGGGCCGGAAGGCCGGGAGGAACCCCTGGGGCGGGCGCACGCT
CGAGTGGATGACCGGCGAGCCCTTGCCGCCCAAGCACAACTTC₅CGGTTCG
CGCCGCTCGTCCGCCGCGGCCCGTACGAGTTCCAGCTCTCCGAGGAGGAC
GCCCCGTGCGACAACCACGCCCGCTGCGTGAGCAGTTCGAAGATCTCGAGA
AGCAGACGCACGCGGCCCGCTCGGGATGTGGTTGTTCTCTCGGGAGCGAG
GTGCTCCTCTTACC₁₀GGGCTCTTCGCGCTGTACGCGGCGTACCGCGAGCT
CTACCCGCGCGATTTCCGCCGAGGCCATCGCGCACAAACGTCGCGATCG
GCACCACCATGACGCTCATCCTGATCGGCAGCAGCTTCACCGTCGCCATG
GCGGTGCACGCCGTC₁₅CGCGCCTCCCACCCGCGGCGCGCGCGCTGTTCTT
CGCGGTGAGCGTGGCGATCGGGATCGTGTTCTCTGTGCTGAAAGGGATCG
AGTACGCGCAGCACTTCCGCGAGGSCATCTTCCCGCGCGCGCCTACCGC
TTCGCGGAGCTCCCGACGTTCCGGCGCGCAGATGGCGTTCACGCTGTACTT
CGCCATGACGGCGCTCCACGCCCTGCACGTCGTGGGCGGGGCGGCCCTCC
TCACGGGGGTGCGGTGGGGGTGCTGGAAGGGCCGGTACTGGGCATACGAC
CAGACGCCCGTGGAGCTCAGCGGCCTTACTGGCACCTCGTGGACATCAT
GTGGATCTTCATCTGGCCGCTCCTCTACCTGACGCGCAAGTGACGGCACA
GCCCCGAGACGACCATGCCGCAAGAGCACGTCGCGGAAAGCACGCCCTGG
ACCCGTTACCTCATGGACGCTGATGSCCCTCATTGCCCTCACGCTCCTGT
CGTTAGCGCTCTCGTTCTTGCGCACGGGGGCTTGGGAAATACCGATCGCG
CTGCTCCTCCCCGTGGTGAAGAGCGTGCTCGGGCTTGCTTTCTTCTTGCT
CCGTGAGGTGCATTAAGGTCATCAAACGCCTTTTCAAACATTGGCTGC
TGGTGTACTTACTTGGACGTTTCATACTGTCTCTTTATGGCCCCCAACG
GTCATTACGTCGACCTACACAATTTCTCTTCCCGCACCCCTCCTACCTGTA
TCTCTAAGCACTGCCTTGCGTCCTGCTCTATTACATTCTACTCCGGCTGT
CCATGTGTGGGATTATATGCGCGAGGTACCTATTCCGCCGTGGAGTCTCC
ATTTACCTCTTGGACCTTGCCCGTCTGAATCTCGATCTCCTCATGCGTT
GGTCCACCATGTATTACCTCCTAGAACTTTATACTCCATATCTCTATATA
TCTAGTTGTGCGTGTAATTGTGTGCATATATTATCGCCACTGTGTATGAA
TACCGTGCCGACGTGCTATATACGAAANTACTCCTCGGTGATATCTCCA
CCTCATATATACCTCCGAGTGTAGTATACGCACGAGTGTATATACTCTTC

5

154

10

CTCTGGTCACGCGACTTCGTGCTGATATGATACCATCGTTCCATGTTACG
CGAAGTTACTCATAAGATCTCCTCACACATCAACGAGTGTACTCCTATGT
GTTTCATACAAACTCGATACCCTTCAGAGTAGTGTGCCTATGTGGTA
TGCATAATGTTAGTATACTTT

5

15

SEQ ID No 79 (>Contig54)

20

10

TGGGAAAGAGGGCCACAGGGGATGTAGCAGGACGCTTAATAGTAAATGAC
GAGGGTGTGCCGACGAGACCCGTAGGAAACAACGGGCACAGACGAGAGCA
ATAAAGGGGGTTTGAAGGTACCCCGGATAGAGTAGAGAAGGCTAGCGGAC
GAGTAAGACGCGGAGGAAATAAGTCGGCGTCGTAGAAGTTCTGTGGAGAA

25

15

GGTACGACTCTTAAAGACCTAGGCGGGAGACAGTTTCCACCCGAGGCAGA
GCAAGACCACAAGATTCAGAGGGAGTAAGGAGTTCGGAATTGGAGAGGTT
GAGGGGCGTGTGAGCCGTCAAGTGGGGCGCGTACGCAAAGAAAGAAGCGT
CCATGTGAGAGGCCAGCGGCCGTTGCGGCCCTACATGGGTGATCACGGT

30

20

GGGCCGTAGGCGGACCGGAGATGAGCGGGCTCCGCCCACCGACGGCGA
GGGGCACGGCGCCTTCGTCCGGGGCGCTCGCGCGCGATCTCGCGCGCGCG
GGGGGCGCCGGCGTCCGCGTGTGCGGACGGCCAGCCGCCCATCGTCCA
CGCCCTCGGGACGTCATCAACGCCGCGCTCCGCAGCCGGGCGGCCTGGA
TGGTCGATCCTGTGCTGATCGACGCGGGCCCTCCACGCAGAGCTTCTCC

35

25

GAGCTCGTCGGCGAGCTCGGGCGCGGCGGTCGACACCTTGATCCTCCT
CGACGTGAACCCCGTGTACGCCGCGCCGGCCGACGTCGATTTGCGGGGCC
TCCTCGCGCGCGTGCCACAGAGCTTGAAGGCCGGGCTCTACGACGACGAG
ACGCCCCGCGCTTGACGCTGGTTCTGTGCCGACCCGGCATTACCTCGAGTC
GTGGGGGACGCGCGGGCGTACGACGGGACGGTCTCGTTCTGTGCAACCCC

40

45

TCGTCCGGCCGCTGTTGACGCGCCGGGCGGTGCCCAGAGCTGCTCGCCGTC
TTCGCGGGGACGAGCGCCCGGATCCCCGGCTGCTGCTGCGCGAGCACTG
GCGCGGCGAGCGCGAGGGGCGGATTTCGAGGCCTTCTGGGGCGAGGCAT
TGAAGCGCGGCTTCTCCCTGACAGCGCCCGGCGAGGCAGACACCGGAG
CTCGCGCCGGCGATCTCGCTCAGGAGCTCGCGCGGCTCGCCGCCGCGCC

50

30

GCGGCCGGCGGCGCGCGCTCGACGTGGCGTTCTCAGGTGCGCGTCGC

55

5

155

10

5

15

20

10

25

15

30

20

35

40

25

45

30

50

55

TCCACGACGGCAGGTTCCGCAACAACCCCTGGCTGCAAGAGCTCCCGCGG
CCGATCACCAGGCTCACCTGGGGCAACGCCGCCATGATGAGCGCGGCGAC
CGCGGCGCGGCTCGGCCTCGAGCGCGGCGATGTCGTCGAGCTCGCGCTGC
GCGGCCGCACGATCGAGATCCCGGCCGTCGTCGTCGCGGGGCACGCCGAC
GACGTGATCAGCGTCGACCTCGGCTATGGGCGCGACGCCGGCGAGGAGGT
CGCGCGCGGGGTGGGCGTGTGGGCGTATCGGATCCGCCCGTCCGACGCGC
GGTGGTTCGCGGGGGGCTCTCCGTGAGGAAGACCGGCGCCACGGCCGCG
CTCGCGCAGGCCCAGCTCGAGCTCTCCAGCACGACCGTCCCATCGCGCT
CAGGAGGACGCTGCCGAGTACCGTGAACAGCCCGGTTTCGCGGAGGAGC
ACAAGGGGCCGCTCCGCTCGATCCTGCCGAGGTCCAGCACACCGGCGCG
CAATGGGCGATGTCCATCGACATGTCGATCTGCACCGGGTGCTCCTCGTG
CGTCGTGGCCTGTGAGGCCGAGAACACGTCCTCGTCGTGCGCAAGGAGG
AGGTGATGCACGGCCGCGAGATGCAGTGGTTGCGGATCGATCAGTACTTC
GAGGGGGGAGGCGACGAGGTGAGCGTCGTCAACCAGCCGATGCTCTGCCA
GCACTGCGAGAAGGCGCCGTGCGAGTACGTCTGTCCGGTGAACGCGACGG
TCCACAGCCCCGACGGCCTCAACGAGATGATCTACAACCGATGCATCGGG
ACGCGCTTTTGTCTCAAACTGCCCGTACAAGATCCGGCGGTTCAATTT
CTTCGACTACAATGCCACGTCCTGTAACGCCGCGCTCCGCAAGCTCC
AGCGCAACCCGACGTGACCGTCCGCGCCCGCGGCGTCATGGAGAAATGC
ACGTACTGCGTGACGCGATCCGAGAGGCGGACATCCGCGCGCAGATCGA
GCGGCGGCCGCTCCGGCCGGGCGAGGTGGTCACCGCCTGCCAGCAGGCCT
GTCCGACCGGCGCGATCCAGTTCGGGTGCTGGATCACGCGGATACCAAG
ATGGTCGCGTGGCGCAGGGAGCCGCGCGGTACGCCGTGCTCCACGACCT
CGGCACCCGGCCCGGACGGAGTACCTCGCCAAGATCGAGAACCCGAACC
CCGAGATTGAATGAGCCATGGCGGGCCCGCTCATCCTGGACGCACCGACC
GACGATCAGCTGTGGAAGCAGCTCCTCGAGCCGGTATGGAAGCCCGCTC
CCGGCTCGGCTGGATGCTCGCGTTCGGGCTCGCGCTCGGCGGCACGGGCC
TGCTCTTCCTCGGATCACCTACACCGTCCTACCGGGATCGGCGTGTGG
GGCAACAACATCCCGGTGCGCTGGGCCTTCGCGATCACCACCTTCGTCTG
GTGGATCGGGATCGGCCACGCCGGGACGTTTCATCTCCGCGATCCTCCTCC

5

10

15

20

25

30

35

40

45

50

55

TGCTCGAGCAGAAGTGGCGGACGAGCATCAACCGCTTCGCCGAGGCGATG
ACGCTCTTCGCGGTCTGTCCAGGCCGCTCTTCCGGTCTCCACCTCGG
CCGCCCCCTGGTTCGCCTACTGGATCTTCCCGTACCCCGCGACGATGCAGG
TGTGGCCGCGAGTTCCGGAGCGCGCTGCCGTGGGACGCCGCCGCGATCGCG
5 ACCTACTTCACGGTGTCTGCTCCTGTTCTGGTACATGGGCCTCGTCCCGGA
TCTGGCGGCGCTGCGCGATCACGCCCGGGCCGCGTCCGGCGGGTGATCT
ACGGGCTCATGTCTGCTCGGCTGGCACGGCGCCGCGACCACTTCCGGCAT
TACCGGTGTGTACGGGCTGCTCCCGGGGCTCGCGACGCCCTCGTCTGT
CTCGGTGCACTCGATCGTGAGCAGCGATTTCCGGATCGCCCTGGTCCCCG
10 GCTGGCACTCGACGCTCTTTCGCGGCTTCTTCTGTCGCGGGCGCGATCTTC
TCCGGGTTTCGCGATGGTGCTCACGCTGCTCATCCCGGTGCGGCGGATCTA
CGGGCTCCATAACGTCTGTACCGCGCGCCACCTCGACGATCTCGGAAGA
TGACGCTCGTGACCGGCTGGATCGTTCATCCTCTCGTACATCATCGAGAAC
25 TTCTTCGCTGGTACAGCGGCTCGGCGTACGAAATGCATCAGTTTTTCA
GACACGCTGCGCGGCCCCGAACAACGCGCCTACTGGGCCAGCACGTCT
15 GCAACGTGCTCGTCATCCAGCTCCTCTGGAGCGAGCGGATCCGGACGAGC
CCCGTTCGCGCTCTGGCTCATCTCCATCCTCGTCAACGTCCGGATGTGGAG
CGAGCGGTTACGCTCATCGTGATGTCTGCTCGAGGAAGAGTTCTCCCGT
CCAAGTGGCACGGCTACAGCCCGACGTGGGTGGACTGGAGCCTCTTCATC
20 GGGTCAGGCGGCTTCTTCATGCTCCTGTTCTGAGCTTTTGGCGGTCTT
TCCGTTTCATCCCCGTGCGGAGGTCAGGAGCTCAACCATGAAGAGCTGG
AGAAGGCTCGGGGCAAGGGGGGGCTGATGGAGACCGGAACGCTCGGCG
AGTTCGACGACCCGAGGCGATGCTCCATGCGATCCGAGAGCTCAGGCGG
40 CGCGGCTACCGCCGGGTGGAAGCGTTTACGCCCCATCCGGTGAAGGGGCT
25 CGACGAGGCGCTCGACCTCCCGCCTCGAACCTCAACCGGATGGTGCTGC
CCTTCGCGATCTGGGGGTCGTGGGCGGCTACTTCGTCCAGTGGTTCTGC
AACGCTTTCCACTATCCGCTGAACGTGGGCGGGCGCCGCTGAACCTCGGC
45 GCCGGCGTTTCATCCCGATCACGTTTCGAGATGGGGGTGCTCTCCACCTCGA
TCTTCGGCGTGCTCATCGGCTTTTACCTGACGAGGCTGCCGAGGCTCTAC
30 CTCCCGCTCTTCGACGCCCCGGGCTTCGAGCGCGTCACGCTGGATCGGGT

5

157

10

5

15

20

10

25

15

30

35

40

25

45

30

50

55

CCTGGTCGGGCTCGACGACACGGAACCTTCCTTCTCGAGCGCCCAGGCGG
AGCGCGATCTCCTCGCGCTCGGCGCCAGGCGCGTCGTCTGGCGAGGAGG
CGCGAGGAGCCATGAGGGCCGGCGCCCCGGCTCGCCCCCTCGGGCGCGCG
CTCGCGCCGTTTCGCCCTCGTCCTGCTCGCCGGGTGCCGCGAGAAGGTGCT
GCCCCGAGCCGGAATTTCGAGCGGATGATCCGCCAGGAGAAATACGGGCTCT
GGGAGCCGTGCGAGCACTTCGACGACGGCCGCGCGATGCAGCACCCGCCC
GAGGGGACCGTCGCGCGCGGGCGCGTCACCGGGCCGCCGGCTATCTCCA
GGGCGTCTTCGACGGGGCGTACGTACCGAGGTGCCGCTCTCGCTCACGG
TCGAGCTCGTGACGCGCGGCCGCGCAGCGCTTCGAGACCTTCTGCGCGCGG
TGCCACGGGATCCTCGGCGACGGCAGCTCGCGCGTGGCGACGAACATGAC
GCTGCGCCCCGCCCGCTCGCTCGTCGACCCGAGGCGCGGAGCTTCCCGC
CGGGCAGGATCTACCAGGTCAATCATCGAGGGCTACGGCTGATGCCGCGC
TACTCGGACGATCTGCCCGACATCGAAGAGCGCTGGGCCGTCTGCGCCTA
CGTGAAGGCGCTTCAGCTGAGCCGCGGAGTGGCCGCGGGCGCCCTCCCGC
CCGCGCTCCGCGGCCGGGCAGAGCAGGAGCTGCGATGAACAGGGATGCCA
TCTAGTACAAGGGCGGCGCGACGATCGCGGCCTCGCTCGCGATCGCGGCG
CTCGSCGCGGTGCGCCGCGATCGTCGCGCGCTTCGTGATCTCCGCCGTT
CTTCTTCTCGTACCTCGCCGCGTGGTCTGTCGCGTATTCTGTCCGTGG
GCGCGCTCGTCACGCTCCTCACCTGCAACGCCATGCGCGCGGGCTGGCCC
ACGGCGGTGCGCCGCCTCCTCGAGACGATGGTGGCGCCGCTGCCCTGCT
CGCGCGCTCTCCGCGCCGATCCTGGTGGCCCTGGACACGCTGTACCCGT
GGATGCACCCGAGCGGATCGCCGGCGAGCACGCGCGGCGCATCCTCGAG
CACAGAACGCCCTACTTCAATCCAGGCTTCTTCGTCTGCGCTCGGCGAT
CTACTTCGCGATCTGGATCGCCGTGCGCCCTCGTGCTCCGCCGGCGATCGT
TCGCGCAGGACCGTGAGCCGAGGGCCGACGTCAAGGACGCGATGTATGGC
CTGAGCGGCGCCATGCTGCCGGTCTGGCGATCACGATCGTCTTCTCGTC
GTTGCACTGGCTCATGTCCCTCGACGCGACCTGGTACTCGACGATGTTCC
CGGTCTACGTGTTTCGCGAGCGCCTTCGTGACCGCCGTCGGCGCGCTCACG
GTCCCTCTCGTATGCCGCGCAGACGTCCGGTTACCTCGCGAGGCTCAACGA
CTCGCACTATTACGCGCTCGGGCGGCTGCTCCTCGCGTTCACGATATTCT

5

10

15

20

25

30

35

40

45

50

55

GGGCCTATGCGGCCTATTTCCAGTTCATGTTGATCTGGATCGCGAACAAG
CCCCACGAGGTCGCCTTCTTCCTCGACCGCTGGGAAGGGCCCTGGCGGCC
GACCTCCGTGCTCGTCGTCTCACGCGGTTCTGTCGTCCCGTTCCTGATCC
TGATGTCGTACGCGATCAAGCGGCGCCCGCGCCAGCTCTCGTGGATGGCG
CTCTGGGTGCTCGCCTCCGGCTACATCGACTTTCCTGGCTCGTGGTGCC
GGCGACAGGGCGCCACGGGTTGCCTTACTGGCTCGACCTCGCGACCC
TGTGCGTCGTGGGCGGCTCTCGACCGGTTCCCGCGTGGCGGCTGCGA
GGGCGGCCGGTGGTCCCGGTCCACGACCCGCGGCTCGAAGAGGCCTTTGC
GTACCGGAGCATATGATGTTCCGTTTCCGTACAGCGAGGTTCCGCCAGGA
GGAGGACACGCTCCCTGGGGGCGCGTGATCTCGCGTTCGCCGTCGTGC
TCGCGATCGGCGGCGCGCTGACGCTCTGGGCTGGCTCGCGATGCGGGCC
CGCGAGGCGGATCTGCGGCCCTCCCTCGCGTTCCTCGAGAAGGATCTCGG
GCCGCGGCGCGAGGTCAGCATGGTCCAGCAGTCGCTGTTCCGACGAGGCGC
GCCTGGGCCAGCAGCTCGTCGAGGCGCAGCGCGGAGCTCCGCCGCTTC
GGCGTCGTGATCGGGAGAGGGGCATCGTGAGCATCCCGATCGACGACGC
GATCGAGCTCATGGTGGCGGAGGGCGCGGATGAGCCGGGCCGTCGCCGT
GGCCCTCCTGCTGGCAGCCGGCCTCGTGTGCGGCCCGGGCGCGCTCCG
AGCCCGTATCGCTTTCGCCCCGCGTGGGCCCGTCCGCGGGCGAGGCCGC
GCTCTGAAACGACGGCTCCGGCGCGGATGAGCGGCCCGAGGCGACCTCCT
GCAACCCACCGCGCTGCGTACAGGGGTAAATCAACTGCATTCCAGGATAC
GGACCGCGCAGCATAAACCTCACCAGCAGACTGAATAATGCCGAACCT
TGAACTTCTATGCCCATGCGTGCGGGATCATCATGCCCAACTATTTAAAC
TGGTCCCTCCGGCAAAGGAACGGACCAGCACCAGAATAAACCCTGTTT
GGCCAGCGCAAAAAATGAAAACTCTTCTCCTTGCGGCTAAAAATAACGA
CTCCGGGAACCGAATGGTATGCAATACACCATGGCAACGCGATTCTGTC
TAACCTTAAACAACTTCCCAGCGACTCGTCCGTGCAACGCTTCTGACGC
AAGACCATGACCCACGAGAACCGGGCGGCGACACTGCCAGTGAAAC
TCGCCTAGGCCCGCCCTGCCTTCACGTATTCACCGTGGGGCGGTCCAAA
ACTAAAAACAATACGTGACTCTTCTCAAATATCGTCGGATAAAGGCCAAC
ACGCGTACCTCCCCCTAAAGGGAAGAAACCCCTACCAGGGTGGACCGTAT

5
10
CCACCCGTGATCCCTGAAACATCTCCAGCCGCGTACAAATTAGGCTTTGA
CAAAACC

SEQ ID No 80 (>Contig55)

5 GGGGAAAGAAGTCGGGAGCAGCAGAAGAGCGCAGCGATTGAAAGCACGAC
15 GCAAGGGCAAAGGACATNAGACGACAACAGAAAGGCGAAGACATGAGAAG
AGGAGAAAGACTCGGAAATGGGCACAAAGCGAGAAAAACGTAACGCTAT
CTGAGAAAGAACACCATAAAGTCAGACTGGGGTAAAGCCATACACGCACC
20 GAAAGCAGACAAGCAAAAGTCATTAACAGATGCAACAAAGAACAAAAGAG
10 AAAGGAAGGACACATGAAGGAGAAAGGCGGCTTCGAGAACTAAAGACGGG
AGCCAACAATATACCTTTCACAGGGGCGAAGAAGGGCCCAAGGTTCAATC
GAAGACGATTGAATCCAAGGAAGTCCAATCGGAAGAAAAAGCATATTGA
25 TAAACCAGACAACGACAGCAGGCCTGAACGTAGGAGCGAGATCGTGAGAC
ATCAGTAGGCAAAACAAGAGCGCTACACCCAGGGTTCGTCAACCTAGAAA
15 GCGCGCTCCTCAAGCCGGTAGCGGCCGCGCGCACCAGCCCGATGCGGGC
GCCTTCGCGCGCGAACC GCCGCACCGTCGCGCGCCCCACGCCCGCTGACG
30 CCCCCGTGATCACCACGACCTCCGGCCGCTGGAATCGCCCATGCCCGTC
GCCTCCCGCTCGCCCGCGGCGTCAAGCAACGTGAATGCCACCTGAGCGT
GTCACTTCCTCAAGCTCGACAGCACGTCCTTGATCCGCTCGGTCCGACCC
20 GCCGTGAGCGACCAGGGGCTCGGCTGCCGCACGACGTGCAGGTGCGCTCG
35 ACGCTCGAGCACCTCGAACC CGTCGTTCCCTGCTCGGTCCGACGAACC
GGAGCGCCACCACGGCCGTGCCACATGGAGGTTGAGAGGGTGATCTCG
GGCAGCCACGCCGCGCAGCGCCGGTCCGACCAGCAGCAGGTGAGCGGCGC
40 GAACGGGTACAGCCCGAGCATCGCCTGGAGCATCGGAACACGCTCGAGC
25 AGGACCAGGCCTGCGGCCAGTTGGCCTTCGGGTACAGCGGGGAACGGG
TGCTCCGCGTCGCGCGGGTGGCCGCTCCAGCACTCGGGCAGCCGGTGGTG
CTCGAACAGCGCCCGCCCTCGAACACGGCGCGGCACAGGAGCGCCACGT
45 GCCCGTGCAGGCCGTACCGCGGAGCCCGAGCGCGATCGCGCCCTGATCG
ACGGGCCAGACCGTCCCCCGGTGATAGCTGTACGGATCGAACGCGGGGTG
30 ATCGGCCGAGAGCGTGCGGATACCCAGCCGGAGAACATGTCCGCCGCGA

50

55

5

160

10

5

15

20

10

25

15

30

35

40

25

45

30

50

55

ACATGCGGATGGCGGTGGGCTCGGCGAGGGCCTGGTCCACGATCCCCGCC
GCGAGGCAGAGCCCCGGATCCGAGCCGATCGAGCGGATCTGGCGCTTGTC
CGGGCCGAGGCCCCATCGCGAAGGTGCGCGCGTCGGGCATCCAGAAGGCGT
CGTTGAACCGCCTCTGCAGCTCGAGCGCCTCGGCGAACAGCCGTCGCGCG
TCGTCTTGCGGGCCGAACCAGAAGAGCAGCTCGGAGAGGCGCAGCTTCGA
CAGGAACACGAAGCCCTGCATCTCGCACGTCCCGATCGGCGGCCGCACCT
GAGAGCCGTCGGCGTGGACGATGGCGTCGTGGAGTCCCTCCAGCCCTGG
TTCTGGATCGACGCGCTCGAGCGGGGCTCGTACTCGTAGAACCCGTCGCC
GTGAGATCGCCCTCCTCGTCGATCCAGCGCATGGCCCTGAGCGCAGGCT
CGATCAGGCGGCCGACGCGCTCGCGATCGCCGGTCCAGTGCCAGAGCTCC
GAGGCGGCCACCGCGTAGAACATCGTCGAGGTGCGCGACGCGTACGTGCG
CCCCAGCGGGTTGTAGTTGAGGTCCGAGAGCGCTCCGTCCCTGGCCTGAT
GCAGCATCCGGTCGGGCTGCTCGTCGCGCCAGTCGTGACGACGCGCCCC
TGCCAGCGCGGGAGCACGAGCGCTGTGCCTGCGAGGATGTGGTTCGTGAG
CGCGCGGCCTGCGTCCCCGCGGCGAGCGGGTCGCGGCCGAAGAGCCCGA
TG TAGATCGGCAGTCCGGCGGCCACCGTCCAGGAGCGCTCGTCTGGTTCG
ATATCGTACATGCGGAGCGCGATGAGATCGCGCTTGCTCGCTCGAGCAC
GGAGAAGACCGTATGCGAGAGCGTGTGCGCGCCCGGGACCGAGAAGGACG
TGGCGCGGTGGTGAAGGTGCCGCGCGCCACGTGCGCGCGTTTCGCGTG
CCGAAGAACGAGCGGCAGCCGGCGAGCAGCGGGAGCGGTTCCGCCCGGAT
CAGGGCGATCACGTGACGCGAGCCGCGCCAGGTGCCATGGGGCTCGAGAT
CGATCGAGAACCAGATCTCGCGGCCAGCGCAGGACGGGGGCGAGCCCGCG
CTCCGGGCCCCGTACGACGATCCCGGCGTCGAAGCGCGCGACGCCGATTTC
TCCGGGGTGTTTCATAGCGGTGCTCGGCGCGGTAGTCGACGCGCAGCGCGC
ATCCGTGCTCGGCGGGCTCGAGCGTCCACCGGGCGTCGCCGCGCTGGAGG
CGCGGGCCGTCCGTCTCCTCCGCGTCGGCGAAGTCGGCGTCGATCTCGAG
CGCCAGCGTGAAGCGCACGCGCTCCTGCGTGAAGCTCGCGACGTCGATGT
CCTCGTGGAAGCCGTCCCCACGAAGCGCGACAGCCGAGCTCGACGGTC
CGCGGCGCCGCGCCGCCCGCTCGGGCGGCGGGCGATGTAGTAGCCGAG
CCAGCTGTCCGGCTCGACGGCGGAGAGCGCGACGGGCGCCGGCGGCCTGC

5

10

15

20

25

30

35

40

45

50

55

CGTCGATCAGGTGGCGGTAGAGGGAGAGGAGGCGGGTGTTCGCGACGAAC
AGGCCGATGTGCGGCTCGGGGGCGATCGAGCCGTCCGGACGCATGCACAG
GACCGTCCGGTTCTGGCTGATGTACAGGGAGCCGGCGGGGGCCTCAGCG
TGGCCAGCGAGCCGAAGGCTCTCTCGAGCGACATGGTCACCTCCAGGCGA
5 GCGCGGGGTGAGGACTGCCAGGGGCGTGCCAACGACGTGGACGCGCTCAC
GCCAAGCCGTGGGCGGGCGGGCGGGCCCCCTGCGCCGAGGCTCAGCCGA
15 GCGCGAGCGGTGTTTCGTGCGCGGAGCGGGCCTCGTCGTGCGCTCCAGG
TAGGCCTTTCCGATGTGCGCCTCGTCCACGAACCCGACGATCGCGCCCTC
GCGCTCGACGACGGGGACCTCGCGGACGCCGTGCGCCACCATCGCCTCGG
20 TCGCCGTCCGAGATCGTCGGTGACCGTCACGGCCACCGCGGCTGCATC
GCGTCGGCGGCCACGGTCATCCGCTCGAGGTCGTGCTCCACCGCGATGAT
CCGGAGCGACTCGGCGGTGATCATGCCGACCATCTTGCGCGACGGTTCGA
25 GCACCGGGAACACCTCCTGCCAGCTCGCGTCGGCCGCCCCCGGAGCATC
TCGCGGGCCGGCGTCCCCGGCACGAACGTACAGAGCGCGCGCCCTCGAT
15 CATGATCTGCCGCACGCGGATGGTCTTGAGCACGTCGAGCGTCGGCACCG
GGTGC GCGGGAGACTCGCGCTGGGTGGGGAGCTGCGCGTGGTAGAGCGAG
30 TGCTTCCGAGCGCGACGAAGGCGACGCCCTCGGCGAGCATCAGCGGAAC
CAGGAGGTCGTAGCTGCCGGCGAGCTCGCAGACCATCAGAGGGAGCTCA
CCGGCACGTGCGCGACGCCCGCTAGAAGGTGCCCATGCCACGAGCGCG
20 AAGCGCCCCGGTTCGATCGCGCGATCGCCGAGCAGGAGCGCCGCGCGCG
35 CCCGAACGCGCCGCCGAAGAGCCCGCCGATGACGAGCGACGGCGCGAAGT
CGCCGGCGCACCCGCCGCTGCCGAGCGTGAGCGACGAGGCGACGATCTTG
GCGCGCAGAGCAGGAGCAGGAGCTCCACGCCGCGCCAGCCCGGTGGAG
40 CCACGTGGCGCCGGTGATCGCGACCTGGACGGCGCCGTACCCGCCGCCGA
25 GCAGCCCGAGCCCTTGCCCGGGGCTCTCGATCCTGCGGCCGACGAACCAG
AGGACGGGCACGCAGAAGAGGGCCAGCGGAGCCCGGAGCCCCGGGCG
CGCCAGGGGGCGATGGG CAGGCGCGCCGCGATCGCCTTCACGCCGCCGA
45 GGCACTTCAGGAAGCCGATCGCGACGAAGGCCAGGAGCAGCGGAGCAGC
GCATAGAGAGGGAGGTGCGACGGGACGAACGCATACTTCGGCGCGTGCGC
30 GAAGAGCGTCGACTCGCCGTAGAACGAGATGAAGACCGAGTAGGAGACCA

5

162

10

CGCTGGCGAGCAGCGCCGGGATCAGCGCCTCGGCCTCGAAGTCGTCCGGG
TAGAGCACCTCGACGGCGAGCAGGGCGGCGCCGAGCGGCGTGCGGAAGAT
GGCGGACATCCCGGCCGCGACCCCCGCGAGCATCAGGATGCGGTGCTCGC
GCCGGCCGACCGCGAGCCCGCGCCCCACGAGCGAGCCGAGCGCTCCTCCG
5 ACCTGCATGGTCGGCCCCCTCGCGACCGCCGGCGCCGCCGAGCCGAGCGT
15 GAGGATCGACGCGACCGCCTTGACCCACGCGACCCGCTTGCGCATCCGGC
CGCCGTGGTGGTGAAGGCCTGGATCATCGCGTCGCCGCCGCCGCCCGCG
GCCTCGGGGGCGAGGCGCCAGGTGAGGATGCCCCCGCCAGCGCTCCGAG
CGCCGGGATCAGCAGCAGCAGCCAGAGCCGACGCTCCGGTGCTCGGGGC
20 CGTCGCCCGCTGAAGATGGCCTCGCCGTGGGCTCGAAGCCGCGCGTAGCCG
10 GCGAGGCGGGCCGAGGAGCAGCTCCTCGACGAGCTCGAGGGCGCCGAAGAA
GAGCACCGCGACGAGGCCCGGATGGCCCCGACGAGCACGGCATGCAGGA
25 TCGTGCGCCCGACGAGCCTGAGATCGAGGGGCGCGACCTCCGAGAGGAGC
GCGGAGAAGGGGCGCCGCCGTGCGACACGCGCGCGCTCTGCTGGGAA
15 TGGTTCCGTCACGGCGATGGT

30

SEQ ID No 81 (>Contig56)

35

GGATCCGGCCGCGAGCGGCTGCGGTGCCGATGCGCTCGACGGTGACGGGC
GGGGTGATCGCGGGTCCGGAGCTCGGTGCGAGCTACTGGGCGGACAACT
20 TCGGCAGCCGGTGCGCTTCGCTGCGGGCGCAAGCGCTGCTGGAGGGTG
GCCCCGCGCTGTTTCATCGAGATGAGCCCGCACCCGATCCTGGTGCCGCCC
CTGGACGAGATCCAGACGGCGGCCGAGCAAGGGGGCGCTGCGGTGGGCTC
GCTGCGGCGAGGGCAGGACGAGCGCGCGACGCTGCTGGAGGCGCTGGGGA
40 CGCTGTGGGCGTCCGGCTATCCGGTGAGCTGGGCTCGGCTGTTCCCCGCG
25 GGCGGCAGGCGGGTTCCGCTGCCGACCTATCCCTGGCAGCACGAGCGGTG
CTGGATCGAGGTCGAGCCTGACGCCCCCGCCTCGCCGAGCCGACCCCA
CCAAGGACTGGTTCTACCGAACGGAAGTGGCCCCGAGGTGCCCCGCGCGCC
45 CCGAAATCGGAGACAGCTCATGGCAGCTGGCTGCTGTTGGCCGACAGGGG
TGGGGTCGGTGAGGCGGTGCGTGCAGCGCTGTCGACGCGCGGACTTTCCT
30 GCACCGTGCTTCATGCGTCGGCTGACGCCTCCACCGTCGCCGAGCAGGTA

50

55

5

10

15

20

25

30

35

40

45

50

55

TCCGAAGCTGCCAGTCGCCGAAACGACTGGCAGGGAGTCCTCTACCTGTG
GGGCCTCGACGCCGTCGTCGATGCTGGGGCATCGGCCGACGAAGTCAGCG
AGGCTACCCGCCGTCGCCACCGCACCCGTCCTTGGGCTGGTTCGATTCTTG
AGCGCTGCGCCCCATCCTCCTCGCTTCTGGGTGGTGACCCGCGGGGCATG
5 CACGGTGGGCGGCGAGCCAGAGGCCTCTCTTTGCCAAGCGGCGTTGTGGG
GCCTCGCGCGCTCGCGGCGCTGGAGCACCCGCTGCCTGGGGTGGCCTC
GTGGACCTGGATCCTCAGAAGAGCCCGACGGAGATCGAGCCCCTGGTGGC
CGAGCTGCTTTCGCCGACCGCGAGGATCAACTGGCGTTCCGCAGCGGTC
GCAGGCACGCAGCACGCCTTGTAGCCGCCCGCCGGAGGGCGACGTCGCA
20 CCGATATCGCTGTCCGCGAGGGGAGCTACCTGGTGACGGGCGGGCTGGG
TGGCCTTGGTCTGCTCGTGGCTCGGTGGCTGGTGAGCGGGGAGCTCGAC
ATCTGGTGCTCACCAGCCGGCACGGGCTGCCAGAGCGACAGGCGTCGGGC
GGAGAGCAGCCGCCGAGGCCCGCGCGCATCGCAGCGGTGAGGGGCT
25 GGAAGCGCAGGGCGCGGGGTGACCGTGGCAGCGGTGGATGTGCGCGAGG
CCGATCCCATGACGGCGCTGCTGGCCGCCATCGAGCCCCCGTTGCGCGGG
GTGGTGACGCGCGCGGCTCTTCCCGTGCGTCACCTGGCGGAGACGGA
CGAGGCCCTGCTGGAGTCGGTGCTCCGTCCCAAGGTGGCCGGGAGCTGGC
TGCTGCACCGGCTGCTGCGCGACCGGCCTCTCGACCTGTTCGTGCTGTTC
TCGTGCGGCGCGGCGGTGTGGGGTGGCAAAGGCCAAGGCGCATACGCCGC
20 GGCCAAATGCGTTCTTCGACGGGCTC3CGCACCATCGCCGCGCGCACTCGC
TGCCGGCGTTGAGCCTCGCCTGGGGCTTATGGGCCGAGGGAGGCATGGTT
GATGCAAAGGCTCATGCAGTCTGAGCGACATCGGGTCTTGCCTATGGC
CACGGGGCCGGCCTTGTGCGGCGCTGGAGCGCCTGGTGAACACCAGCGCTG
40 TCCAGCGTTCGGTCACACGGATGGACTGGGCGCGCTTCGCGCCGGTCTAT
GCCGCGCAGGGCGGCGCAACTTGCTTTCGGCTCTGGTCGCGGAGGACGA
25 GCGCGCTGCGTCTCCCCGGTGCCGACGGCAAACCGGATCTGGCGCGGCC
TGTCGGTTGCGGAGAGCCGCTCAGCCCTCTACGAGCTCGTTCGCGGCATC
GTGCCCCGGTGCTGGGCTTCTCCGACCCGGGCGCGCTCGACGTCGGCCG
45 AGGCTTCGCCGAGCAGGGGCTCGACTCCCTGATGGCTCTGGAGATCCGTA
30 ACCGCCTTCAGCGCGAGCTGGGCGAACGGCTGTGCGCGACTCTGGCCTTC

5

164

10

15

20

25

30

35

40

45

50

55

GACCACCCGACGGTGGAGCGGCTGCTGGCGCATCTCCTCACCACGTGCT
GAAGCTGGAGGACCGGAGCGACACCCGGCACATCCGGTCGGTGGCGGCGG
ATGACGACATCGCCATCCTCGGTGCCGCTGCCGGTCCCAGGTGGGGAT
GAGGGCCTGGAGACATACTGGCGGCATCTGGCCGAGGGCATGGTGGTCAG
5 CACCGAGGTGCCAGCCGACCGGTGSCGCGCGGCGGACTGGTACGACCCCG
15 ATCCGAGGTTCCGGGCCGGACCTATGTGGCCAAGGGTGCTTCTCTCCGC
GATGTGCGCAGCTTGGATGCGGCGTTCTTCGCCATTTCCCCTCGTGAGGC
GATGAGCCTGGACCCGCAACAGCGCTGTGCTGGAGGTGAGCTGGGAGG
CGATCGAGCGCGCTGGCCAGGACCGATGGCGCTGCGCGAGAGCGCCACG
20 10 GCGGTGTTCTGTGGGCATGATCGGGAGCGAGCACGCCGAGCGGGTGCAGGG
CCTCGACGACGACGCGGCGTTGCTGTACGGCACCACCGCAACCTGCTCA
GCGTCGCCGCTGGACGGCTGTCTCTCTCTGGGTCTGCACGGCCCCGACG
25 ATGACGGTGGACACCGCCTGCTCGTCTGCTGGTGGCGTTGCACCTCGC
CTGCCAGAGCCTGCGATTGGGCGAGTGGCAGCAGGCCCTGGCCGGCGGGT
15 CCAGCGTGCTTTTGTGCGCCGCGGTCAATTCGTGCGGCGCTGCGCATGCGT
TTGCTTTCGCCAGATGGGCGGTGCAAGACGTTCTCGGCCGCTGCAGACGG
30 CTTTGC CGGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGCTCCGTG
ACGCGCAGCGCGACCGCGACCCCATCTCTGGCGGTGGTCAGGAGCACGGCG
ATCAACCACGATGGCCCCGAGCAGCGGGCTCACGGTGGCCAGCGGTCCTGC
20 CCAGCAGGCGTTGCTACGCCAGGCGCTGGCGCAAGCGGGCGTGGCGCCGG
35 CCGAGGTGATTTCTGTGGAGTGCCACGGGACGGGACAGCGCTGGGTGAC
CCGATCGAGGTGCAGGCGCTGGGCGCGGTGTACGGGCGGGGCGCCCCGC
GGAGCGGCCGCTCTGGCTGGGCGCTGTCAAGGCCAACCTCGGCCACCTGG
40 AGGCCGCGGGCGGGCTTGGCCGGCGTGCTCAAGGTGCTCTTGGCGCTGGAG
25 CACGAGCAGATTCCGGCTCAACCGAGCTCGACGAGCTCAACCCGCACAT
CCCGTGGGCAGAGCTGCCAGTGGCGGTGTCCGAGGGCGGTCCCCTGGC
CGCGCGGCGCGCCCCGCGTCTGTGAGGCGTGAGCGCTTTCGGCCTGAGC
45 GGGACCAACGCGCATGTGGTGTGAGGAGGCGCCGGCGGTGGAGCCTGT
GGCCGCGGCCCCCGAGCGCGCAGCGAGCTGTTCTCTGTCTGGCGAAGA
30 GCGCGCGGCGCTGGATGCGCAGCGAGCCGGCTGCGGGACCACCTGGAG

5

165

10

5

15

20

10

25

15

30

35

40

25

45

50

55

AAGCATGTCCAGCTTGGCCTCGGCGATGTGGCGTTCAGCCTGGCGACGAC
GCGCAGCGCGATGGAGCACCGGCTGGCGGTGGCCGCGAGCTCGCGCGAGG
CGCTGCGAGGGGCGCTTTCGGCCGCGAGCGAGGGGCACACGCCGCCGGGA
GCCGTGCGTGGGCGGCCCTCGGGCGGCAGCGCGCCGAAGGTGGTCTTCGT
GTTTCCCGCCAGGGCTCGCAGTGGGTGGGCATGGGCCGAAAGCTCATGG
CCGAAGAGCCGGTCTTCGGGGCGGCGCTGGAGGGTTGCGACCGGGCCATC
GAGGCGGAAGCGGGCTGGTCTGCTCGGGGAGCTCTCCGCCGACGAGGC
CGCCTCGCAGCTCGGGCGCATCGACGTGGTTAGCCGGTGTGTTCGCCA
TGGAAGTAGCGCTTCTGCGCTGTGGCGGTCTGCGGGAGTGGAGCCGAA
GCGGTGGTGGGCCACAGCATGGGCGAGGTTGCGGCGGCGACGTGGCCGG
CGCGCTGTGCTCGAGGACGCGGTGGCGATCATCTGCCGGCGCAGCCGGC
TGCTGCGGCGGATCAGCGGTACAGGGGAGATGGCGCTGGTTCGAGCTGTCTG
CTGGAGGAGGCCGAGGCGGCGCTGCGTGCCATGAGGGTCTGGCTGAGCGT
GGCGGTGAGCAACAGCCCGCGCTCGACCGTGTCTGCCGGCGAGCCGGCGG
CGCTCTCGGAGGTGCTGGCGGCGCTGACGGCCAAGGGGTGTTCTGGCGG
CAGGTGAAGGTGAGCGTCGCCAGCCATAGCCCGCAGGTGACCCGCTGCG
CGAAGAGCTGATCGCGGCGCTGGGAGCGATCCGGCCGCGAGCGGCTGCGG
TGCCGATGCGCTCGACGGTGACGGGCGGGGTGATCGCGGGTCCGGAGCTC
GGTGCGAGCTACTGGGCGGACAACCTTCGGCAGCCGGTGCCTTCGCTGC
GGCGGCGCAAGCGCTGCTGGAGGGTGGCCCCGCGCTGTTTCATCGAGATGA
GCCCCACCCGATCCTGGTGCCGCCCTGGACGAGATCCAGACGGCGGCC
GAGCAAGGGGCGCTGCGGTGGGCTCGCTGCGGCGAGGGCAGGACGAGCG
CGCGACGCTGCTGGAGGCGCTGGGGACGCTGTGGGCGTCCGGCTATCCGG
TGAGCTGGGCTCGGCTGTTCCCCGCGGGCGGAGGCGGGTTCGCTGCCG
ACCTATCCCTGGCAGCACGAGCGGTACTGGATCGAGGACAGCGTGCATGG
GTCAAGCCCTCGCTGCGGCTTCGGCAGCTTCGCAACGGCGCCACGGACC
ATCCGCTGCTCGGGGCTCCATTGCTCGTCTCGGCGCGACCCGGAGCTCAC
TTGTGGGAGCAAGCGCTGAGCGACGAGAGGCTATCCTACCTTTCGGAACA
TAGGGTCCATGGCGAAGCCGTGTTGCCCGAGCGGCGTATGTAGAGATGG
CGCTCGCCCGGCGTAGATCTCTATGGCACGCGACGCTGGTGCTGGAG

5

10

15

20

25

30

35

40

45

50

55

CAGCTGGCGCTCGAGCGAGCCCTCGCCGTGCCCTCCGAAGCGGACGCAT
CGTGCAAGTGGCCCTCAGCGAAGAAGGTCCCGGTGCGGCCTCATTCCAGG
TATCGAGTCGTGAGGAGGCAGGTAGGAGCTGGGTGCGGCACGCCACGGGG
CACGTGTGTAGCGGCCAGAGCTCAGCGGTGGGAGCGTTGAAGGAAGCTCC
GTGGGAGATTCAACGGCGATGTCCGAGCGTCCTGTCGTGCGAGGCGCTCT
ATCCGTGCTCAACGAGCACGCCCTCGACTATGGTCCCTGCTTCCAGGGC
GTGGAGCACGTGTGGCTCGGCACGGGGGAGGTGCTCGGCCGGGTACGCTT
GCCAGGAGACATGGCATCCTCAAGTGGCGCCTACCGGATTCATCCCGCCT
TGTTGGATGCATGTTTTAGGTGCTGACAGCGCTGCTCACCACGCCGAA
TCCATCGAGATTCGGAGGCGGCTGACGGATCTCCACGAACCGGATCTCCC
GCGGTCCAGGGCTCCGCTGAATCAAGCGGTGAGTGACACCTGGCTGTGGG
ACGCCGCGCTGGACGGTGGACGGCGCCAGAGCGCGAGCGTGCCCGTCGAC
CTGGTGCTCGGCAGCTTCCATGCGAAGTGGGAGGTATGGAGCGCCTCGC
GCAGGCGTACATCATCGGCACCTCTCCGCATATGGAACGTCTTCTGCGCTG
CTGGAGAGCGTCACACGATAGACGAGTTGCTCGTCAGGCTTCAAATCTCT
GTCGTCTACAGGAAGGTCATCAAGCGATGGATGGAACACCTTGTCGCGAT
CGGCATCCTTGTAGGGGACGGAGAGCATTTTGTGAGCTCTCAGCCGCTGC
CGGAGCCTGATTTGGCGGCGGTGCTCGAGGAGGCCGGGAGGGTGTTGCC
GACCTCCCAGTCCTATTTGAGTGGTGCAAGTTTGCCGGGGAACGGCTCGC
GGACGTATTGACCGGTAAGACGCTCGCGCTCGAGATCCTCTTCCCTGGTG
GCTCGTTTCGATATGGCGGAGCGAATCTATCGAGATTCGCCCATCGCCCGT
TACTCGAACGGCATCGTGCGCGGTGTCGTGAGTCGGCGGCGCGGGTGGT
AGCACCGTCGGGAATGTTTACGATCCTTGAGATCGGAGCAGGGACGGGCG
CGACCACCGCCCGCTCCTCCCGGTGTTGCTGCCTGACCGGACGGAGTAC
CATTTACCGATGTTTCTCCGCTCTTCTTGCTCGCGCGGAGCAAAGATT
TCGAGATTATCCATTCTGAAGTATGGCATTCTGGATGTCGACCAGGAGC
CAGCTGGCCAGGGATACGCACATCAGAGGTTTGACGTCATCGTCGGGCC
AATGTCATCCATGCGACCCGCGATATAAGAGCCACGGCGAAGCGTCTCCT
GTCGTTGCTCGCGCCCGGAGGCCTTCTGGTGCTGGTCGAGGGCACAGGGC
ATCCGATCTGGTTCGATATCACCACGGGATTGATTGAGGGGTGGCAGAAG

5

10

15

20

25

30

35

40

45

50

55

TACGAAGATGATCTTCGTATCGACCATCCGCTCCTGCCTGCTCGGACCTG
GTGTGACGTCCTGCGCCGGGTAGGCTTTGCGGACGCCGTGAGTCTGCCAG
GCGACGGATCTCCGGCGGGGATCCTCGGACAGCACGTGATCCTCTCGCGC
GCGCCGGGCATAGCAGGAGCCGCTTGTGACAGCTCCGGTGAGTCGGCGAC
CGAATCGCCGGCCGCGCGTGCAGTACGGCAGGAATGGGCCGATGGCTCCG
CTGACGTGCTCCATCGGATGGCGTTGGAGAGGATGTACTTCCACCGCCGG
CCGGGCCGGCAGGTTTGGGTCCACGGTCGATTGCGTACCGGTGGAGGCGC
GTTACGAAGGCGCTCGCTGGAGATCTGCTCCTGTTTGAAGACACCGGGC
AGGTCGTGGCAGAGGTTTCAAGGGGCTCCGCCTGCCGACGCTCGAGGCTTCT
GCTTTCGCGCCGCGGGACCCGCGGGAAGAGTGGTTGTACGCTTTGGAATG
GCAGCGCAAAGACCCTATACCAGAGGCTCCGGCAGCCGCGTCTTCTTCCT
CCGCGGGGGCTTGGCTCGTGCTGATGGACCAGGGCGGGACAGGCGCTGCG
CTCGTATCGCTGCTGGAAGGGCGAGGCGAGGCGTGCCTGCGCGTCATCGC
GGGTACGGCATACGCCTGCCTCGCGCCGGGGCTGTATCAAGTCGATCCGG
CGCAGCCAGATGGCTTTCATACCCTGCTCCGCGATGCATTGGGCGAGGAC
CGGATTGTGCGCGCGGTAGTGATATGTGGAGCCTTGATGCGACGGCAGC
AGGGGAGAGGGCGACAGCGGAGTCGCTTCAGGCCGATCAACTCCTGGGGA
GCCTGAGCGCGCTTTCTCTGGTGCAGGCGCTGGTGCGCCGGAGGTGGCGC
AACATGCCGCGGCTTTGGCTCTTGACCCGCGCCGTGCATGCGGTGGGCGC
GGAGGACGCAGCGCCTCGGTGGCCGAGGCGCCGGTGTGGGGCCTCGGTC
GGACGCTCGCGCTCGAGCATCCAGAGCTGCGGTGCACGCTCGTGGACGTG
AACC CGGCGCCGTCTCCAGAGGACGAGCCGCACTGGCGGTGGAGCTCGG
GGCAGCGACAGAGAGGACCAGGTGCGATTGCGCTCGGATGGCCGCTACG
TGGCGCGCCTCGTGCGGAGCTCCTTTCCGGCAAGCCTGCTACGGATTGC
GGCATCCGGGCGGACGGCAGCTATGTGATCACCATGGCATGGGGAGAGT
GGGGCTCTCGGTGCGCAATGGATGGTATGCAGGGGGCCCGCCATGTGG
TGCTCGTGGATCGCGGCGGCGCTTCCGAGGCATCCCGGGATGCCCTCCGG
TCCATGGCCGAGGCTGGCGCGGAGGTGCAGATCGTGGAGGCCGACGTGGC
TCGGCGCGACGATGTCGCTCGGCTCCTCTCGAAGATCGAACCGTCCGATGC
CGCCGCTTCCGGGGATCGTGTACGTCGACGGGACCTTCCAGGGCGACTCC

5

10

15

20

25

30

35

40

45

50

55

TCGATGCTGGAGCTGGATGCCCGTCGCTTCAAGGAGTGGATGTATCCCAA
GGTGCTCGGAGCGTGGAACCTGCACGCGCTGACCAGGGATAGATCGCTGG
ACTTCTTCGTCTGTATTCTCTCGGSCACCTCGCTTCTGGGCTTGCCAGGA
CAGGGGAGCCGCGCCCGCGGTGACGCCTTCTTGGACGCCATCGCGCATCA
CCGGTGCAAGGTGGGCCTTACAGCGATGAGCATCAACTGGGGATTGCTCT
CCGAAGCATCATCGCCGCGACCCCGAACGACGGCGGAGCACGGCTCGAA
TACCGGGGGATGGAAGGCCTCACGCTGGAGCAGGGAGCGGCGGCGCTCGG
GCGCTTGCTCGCACGACCCAGGGCGCAGGTAGGGGTGATGCGGCTGAATC
TGCGCCAGTGGTTGGAGTXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXTATGGTATAACTTATTGATTATAATACAGTATACAAAGGTCCCTTT
TCAGSGACCCTTTCGTATGTTGTAGCTGATTTTATTTCTTCTTTCTTT
TGGGTGCTTTTAATAGCCATATAATGATGTTTCATATAAAGTTAAAAGTTT
CCCCATCTGTCTATCTATAGCGTCATGTTTTTCAGGATTTCTTAATTTCT
GCAGCAGTTTAGATATATATTTTCGCAGGTAACCTTATTATTAATGAATTA
TGGGATTGATACATTTCCCTTTGTCCAATCTAAATCTTTCTGATTTCAA
TACCTCACTCTCATCTCTATTAATTATAAAAACCGAGCTTTCCATACGGAC
CTGTAAATATGATTGTAATTGTCTATACTCAGATGGACCTAATTTCTTCA
AAATTTTATAGCATCAAAAACAACCTTGCTTGTTTTATAGTCCCTCCAATAC
TCGTTTCCAAAAATCAGATTTGCCACCATTGGTGCCTATAATATCTCGTC
TCTGAACCTGCGTTACCATTTGGATGGGACTTGATGTCTGTTAGGTGGGAT
GCAAATACTATTCTTAGTGCGTCTAAACACCATTGTTCAAATTCAGTGGC
ACCTTCATTTCCCTATCGGTATCTGATCTAGATGAGTAGTTATTTGACCTA
TTGTTTTATTTCTAATGGCTGAATTATCTGAAATAATATTAATATCATAT
TCGTCATTTATCTCCTCTGCTTCCCTCTGGAGCAAGTGCATTACGATTTAG
ATTCAAACCAAGCCAGTAACATGGATGAATTAATAATTTCTCATTACTTT
CAAAACCTTTATCTGGAGTTTCGCGCGTCATGGCAAAATGAATAGGATGAG
GTGTTTTATCACGAATACCAACAAATCCTACACTATACAACTTTGGAG
AATCCCACTTGCCCTTTAACAACCTGATTTTCAGACGTTATTTTAGGATCTC
CATTTCTTCGATTAATTCGAAAGATGCTTCTATTTTTTTTAAGCACGTA
TAAACTGTTAATTCAGGTTCAATGCTACGAAATGCACTAGTTATAACCTG

5

169

10

5

15

20

10

25

15

30

35

20

40

25

45

30

50

55

TATTGAAGGAAAGATCTTCTGATACTCTTTCCAGAGATCTTCAAGTCTGG
CCATGGAAATTGACTTGGCTGCATATTCTAGGTCAGTGTATGATAGTT
TCTCTATTCTCTCTGAATGCGGAAAAAAGCTTCATTCAACAATGATAG
TAAATCCCTGGGCCGTAAAGGGTAAATTGCAAACATCGCTTAAACCAT
TCCTCCCTTTAAGATCATCCGCTGTGCATCTATCCCAAACCTCGTTGATCT
TTCTCAATATCTAGCTTAAATGCTACTTTTCATTCTTTTAGCTGACAGCAT
TAGGAGTTGTGCCCAGTCCCAATGCAACCTTATGACTTGACCCTCTATAT
TTCTCGAGTAATCAGGATCTTCCTTTGATAGCGACCTAAATATATTATCC
CTTAAAAAATTATTTGGACGAATGCATTTTGCTTTTGTATTTAATTCAAT
AGATGCATATGCTAGACCTGCAATGATTCCAATTCCTATATTATCCGGTT
CATACGCCTCATCTAGCTTATCCATTAATATGACAACTTTCTGTCTGAG
CGTTCAAGAAGTGATACTATATTATTTTCTATTTCTGAGATATTCAAATT
GAATTGAAGATCACCAATTGATTCTTCTGGGTTATTTTCATCTAAATACT
CCTTTGCGACAAGCCTACACTTTCTTAAATGTCACCTTGTGCAGAATTC
CATTTTTTCAAATGTTCAATTCAACAATGTTTCTGATGATATTTGAGATGA
CAATTTGTAATGAGATGATATATATGATGCTATCTCCATTAGCATAGCGT
ATCGCCATAATAGTCTTGTTGCTGCCCTTGCTAAATTAATGATCCTGTA
AATGGTTTCAACATTGATCTGAAACCAATAATTGAGAATCGTCTGGTGA
GAAACTCAGGATTAATATTTTTTGTCTTTCTTCCAATGCTCATTTAGCT
GAATAAATAAAGCACTTTTACCTGTCCCTCGTCTACCAACAACAATGGTC
CTGTCATCAGTTTCAATTAGAGTCCTAAAGTCAGCAGTTTCAATGAAAGC
ATTACTCAACATCTTTTTATCATTTTCTGCTGTCGTATCACCAAACGGAT
TAGACTTTGAAGTAATATTCAATTCATATTCAACCTTTTATGTTAGTTG
CTTTCATTTATTACTTTATATACTGTTGAACGAGCAATATTCATTGTTTT
TGATATATGTGAGGCACCTAACCCTGTTGCCACATATTTAATACTGCAT
CTCTATCTATTTTCTTTTTCTACCAAAAACAACCTCTTTTGCCA

SEQ ID No 82 (>Contig57)

TCATCTATTGTATAGTTTGTATATTGATATGATATAATTATAACATATAA
ACAGTAACTTTCTCTACGTAGATCGAGGAGAAGACTCAATTTGTTGACA

5

170

10

15

20

25

30

35

40

45

50

55

TCXXXXXXXXXXXXXXXXXXXXXXXXXXXXACTACTCGCATACCGTTG
CGCAACAGCGGCGCAGGAGCAGGACGCATACGACATCACCGGCAATACG
CTCAGCGTCGCCGACCGACGGTTGTCTTATACGCTAGGGCTGCAGGGACC
CTGCCTGACCGTCGACACGGTCTGCTCGTCGTCGCTCGTGGCCATCCACC
5 TTGCCTGCCGAGCCTGCGCGCTCGCGAGAGCGATCTCGCGCTGGCGGGA
GGCGTCAACATGCTCCTTTTCGTCCAAGACGATGATAATGCTGGGGCGCAT
CCAGGCGCTGTCGCCGATGGCCACTGCCGGACATTGACGCCTCGGCCA
ACGGGTTTCGTCCGTGGGGAGGGCTGCGGTATGGTCGTGCTCAAACGGCTC
TCCGACGCCCAGCGACACGCGATCGGATCTGGGCTCTGATCCGGGGTTTC
10 GGCCATGAATCAGGATGGCCGGTTCGACAGGGTTGATGGCACCCAATGTGC
TCGCTCAGGAGGCGCTCTTGCGCGAGGCGCTGCAGAGCGCTCGCGTCGAC
GCCGGGGCCATCGGTTATGTGAGACCCACGGAACGGGGACCTCGCTCGG
CGACCCGATCGAGGTCGAGGCGCTGCGTGCCGTGTTGGGGCCGGCGCGGG
CCGATGGGAGCCGCTGCGTGCTGGGCGCAGTGAAGACAAACCTCGGCCAC
15 CTGGAGGGCGCTGCAGGCGTGGCGGGTTTGATCAAGGCGGCGCTGGCTCT
GCACCACGAACTGATCCCGCGAAACCTCCATTTCCACACGCTCAATCCGC
GGATCCGGATCGAGGGGACCGCGCTCGCGCTGGCGACGGAGCCGGTGCCG
TGGCCGCGGGCGGGCCGACCGCGCTTCGCGGGGGTGAGCGGTTCCGGCCT
CAGCGGCACCAACGTCCATGTCGTGCTGGAGGAGGCGCCGGCCACGGTGC
20 TCGCACCGGCGACGCCGGGGCGCTCAGCGGAGCTTTTGGTGCTGTCGGCG
AAGAGCGCCGCGCGCTGGACGCACAGGCGGCGCGGCTCTCAGCGCACAT
CGCCCGGTACCCGGAGCAGGGTCTCGGAGACGTGCGGTTACGCCTGGTAT
CGACGCGTAGCCCGATGGAGCACCGGCTCGCGGTGGCGGCGACCTCGCGC
40 GAGGCGCTGCGAAGCGCGCTGGAGGTTGCGGCGCAGGGGACACCCCGGC
25 AGGCGCGGCGCGCGGAGGGCCGCTTCCTCGCCCGCAAGCTCGCCTTCC
TGTTCCCGGGCAGGGCGCGCAGGTGCCGGGCATGGGCCGTGGGTTGTGG
GAGGCGTGGCCGGCGTTCCGCGAGACCTTCGACCGGTGCGTCACGCTCTT
CGACCGGGAGCTCCATCAGCCGCTCTGCGAGGTGATGTGGGCCGAGCCGG
GCAGCAGCAGGTGCTCGTTGCTGGACCAGACGGCGTTACCCAGCCGGCG
30 CTCTTTGCGCTGGAGTACGCGCTGGCCGCGCTCTTCGGTTCGTGGGGCGT

5

171

10

15

20

25

30

35

40

45

50

55

GGAGCCGGAGCTCGTCGCTGGCCATAGCCTCGGCGAGCTGGTGGCCGCT
GCGTGGCGGGTGTGTCTCCCTCGAGGACGCCGTGCGCTTGGTGGTCGCG
CGCGGCCGGTTGATGCAGGCGCTGCCGCCGGCGGCGCGATGGTATCGAT
CGCCGCGCCGGAGGCCGACGTGGCTGCCGCGGTGGCGCCGCACGCAGCGT
5 TGGTGTGATCGCGGCAGTCAATGGGCCGGAGCAGGTGGTGTATCGCGGGC
GCCGAGAAATTCTGTGCAGCAGATCGCGCGGGCGTTGCGGGCGCGGGGGC
GCGAACCAAACCGCTGCATGTCTCGCACGCGTTCCTACTCGCCGCTCATGG
ATCCGATGCTGGAGGCGTTCCGGCGGGTGACTGAGTCGGTGACGTACCGG
CGGCCTTCGATCGCGCTGGTGAGCAACCTGAGCGGGAAGCCCTGCACCGA
10 TGAGGTGAGCGCGCCGGTTACTGGTGCGTCACGCGCGAGAGGCGGTGC
GCTTCGCGGACGGAGTGAAGGCGCTGCACGCGGCCGGTGCGGGCCTCTTC
GTCGAGGTGGGGCCGAAGCCGACGCTGCTCGGCCTTGTGCCGGCCTGCCT
GCCGGATGCCAGGCCGGTGCTGCTCCAGCGTCGCGCGCCGGGCGTGACG
25 AGGCTGCGAGCGCGCTAGAGGCGCTGGGTGGGTTCGGGTGCTCGGTGGA
TCGGTCACCTGGTCGGGTGTCTTCCTTCGGGCGGACGGCGGTACCGCT
GCCAACCTATCCCTGGCAGCGCGAGCGTTACTGGATCGAAGCGCCGGTCG
30 ATCGTGAGGCGGACGGCACCGGCCGTGCTCGGGCGGGGGCCACCCCTT
CTGGGTGAAGTCTTTTCCGTGTCGACCCATGCCGGTCTGCGCCTGTGGGA
GACGACGCTGGACCGAAAGCGGCTGCCGTGGCTCGGCGAGCACCGGGCGC
20 AGGGGGAGGTGCTGTTTCTGGCGCCGGGTACCTGGAGATGGCGCTGTCG
TCGGGGGCCGAGATCTTGGGCGATGGACCGATCCAGGTACCGGATGTGGT
GCTCATCGAGACGCTGACCTTCGCGGGCGATACGGCGGTACCGGTCCAGG
TGGTGACGACCGAGGAGCGACCGGGACGGCTGCGGTTCCAGGTAGCGAGT
40 CGGGAGCCGGGGGAACGTCGCGCGCCCTTCGGATCCACGCCCGCGGCGT
25 GCTGCGCCGGATCGGGCGCGTCGAGACCCGGCGAGGTGGAACCTCGCCG
CCCTGCGCGCCCGGCTTCATGCCGCCGTGCCCGCTGCGGCTATCTATGGT
GCGCTCGCCGAGATGGGGCTTCAATACGGCCCCGGCGTTGCGGGGGCTCGC
45 CGAGCTGTGGCGGGGTGAGGGCGAAGCGCTGGGCGAGGTGAGACTGCCTG
AGGCCGCCGGCTCCGCGACAGCCTACCAGCTGCATCCGGTGTGCTGGAC
30 GCGTGCGTCCAAATGATTGTTGGCGCGTTCCGCCATCGCGATGAGGCGAC

5

172

10

5

15

20

10

25

15

30

35

20

40

25

45

30

50

55

GCCGTGGGCGCCGGTGGAGGTGGGCTCGGTGCGGCTGTTCCAGCGGTCTC
CTGGGGAGCTATGGTGCCATGCGCGCGTCGTGAGCGATGGTCAACAGGCC
TCCAGCCGGTGGAGCGCCGACTTTGAGTTGATGGACGGTACGGGCGCGGT
GGTCGCCGAGATCTCCCGGCTGCTGGTGGAGCGGCTTGGAGCGGTGTAC
GCCGGCGGACGACGAGCCGACTGGTTCCTGGAGCTGGATTGGGAGCCCGCG
GCGCTCGGTGGGCCCCAAGATCACAGCCGGCCGGTGGCTGCTGCTCGGCGA
GGGTGGTGGGCTCGGGCGCTCGTTGTGCTCGGCGCTGAAGGCCCGCGGCC
ATGTCGTGCTCCACGCCCGGGGGACGACAGGACTGCAGGAATGCGC
GCGCTCCTGGCCAACCGTTTCGACGGCCAGGCCCGACGGCCGTGGTGCA
CCTCAGCAGCCTCGACGGGGCGGCCAGCTCGGCCCGGGGCTCGGGGCGC
AGGGCGGCTCGACGCGCCCCGAGCCAGATGTCGATGCCGATGCCCTC
GAATCGGCGCTGATGCGTGGTTGCGACAGCGTGCTCTCCCTGGTGCAAGC
GCTGGTCGGCATGGACCTCCGAAACGCGCCGCGGCTGTGGCTCTTGACCC
GCGGGGCTCAGGCGGCCCGCCGGCGATGTCTCCGTGGTGCAAGCGCCG
CTGTGGGGCTGGGCGCACCATCGCCTTGAGCACGCCGAGCTGCGCTG
TATCAGCGTCGACCTCGATCCAGCCGAGCCTGAAGGGGAAGCCGATGCTT
TGCTGGCCGAGCTACTTGACAGATGATGCCGAGGAGGAGGTGCGCTGCGC
GGTGGCGACCGGCTCGTTGCGCGGCTCGTCCACCGGCTGCCCCAGCTCA
GCGCCGGGAGAAGGTGAGCCCCCGGTGACAGGCCGTTCCGGCTAGAGA
TCGATGAACCCGGCGCGCTGGACCAACTGGTGCTCCGAGCCACGGGGCGG
CGCGCTCCTGGTCCGGGCGAGGTGAGATCTCCGTGGAAGCGGCGGGGCT
CGACTCCATCGACATCCAGCTGGCGTTGGGCGTTGCTCCCAATGATCTGC
CTGGAGAAGAAATCGAGCCGTTGGTGCTCGGAAGCGAGTGCGCCGGGCGC
ATCSTCGCTGTGGGCGAGGGCGTGAACGGCCTTGTGGTGGGCCAGCCGGT
GATCGCCCTTGCGGCGGGAGTATTGCTACCCATGTACACAGTCGGCCA
CGCTGGTGTGCTCGGCCTCTGGGGCTCTCGGCGACCGAGGCGGCCGCG
ATGCCCCCTCGCGTATTTGACGGCCTGGTACGCCCTCGACAAGGTCCGCCA
CCTGCAGGCGGGGAGCGGGTGTGATCCATGCGGAGGCCGGTGGTGTG
GTCTTTGCGCGGTGCGATGGGCGCAGCGGTGGGCGCCGAGGTGTATGCG
ACC3CCGACACGCCCCGAGAACCGTGCCTACCTGGAGTCGCTGGGCGTGCG

5

10

15

20

25

30

35

40

45

50

55

GTACGTGAGCGATTCCCGCTCGGGCCGGTTCGTACAGACGTGCATGCAT
GGACGGACGGCGAGGGTGTGGACGTCTGTGCTCGACTCGCTTTCGGGCGAG
CGCATCGACAAGAGCCTCATGGTCCTGCGCGCCTGTGGTCGCCTTGTGAA
GCTGGGCAGGCGCGACGACTGCGCCGACACGCAGCCTGGGCTGCCGCCGC
5 TCCTACGGAATTTTTCCTTCTCGCAGGTGGACTTGCGGGGAATGATGCTC
GATCAACCGCGAGGATCCGTGCGCTCCTCGACGAGCTGTTGCGGTTGGT
15 CGCAGCCGGTGCCATCAGCCCACTGGGGTTCGGGTTGCGCGTTGGCGGAT
CCCTCACGCCACCGCCGGTCGAGACCTTCCCGATCTCTCGCGCAGCCGAG
GCATTCCGGAGGATGGCGCAAGGACAGCATCTCGGGAAGCTCGTGCTCAC
20 GCTGGACGACCCGGAGGTGCGGATCCGCGCTCCGGCCGAATCCAGCGTCG
10 CCGTCCGCGCGGACGGCACCTACCTTGTGACCGGCGGTCTGGGTGGCCTC
GGTCTGCGCGTGSCCGGATGGCTGGCCGAGCGGGGCGCGGGCAACTGGT
GCTGGTGGGCCGCTCCGGTGCGGCGAGCGCAGAGCAGCGAGCCGCCGTGG
25 CGGCGCTGGAGGCCACCGCGCGCGCTCACGGTGGCGAAAGCGGACGTC
15 GCCGATCGGTACAGATCGAGCGGGTCTCCGCGAGGTTACCGCCTCGGG
GATGCCGCTGCGGGGTGTCGTGCATGCGGCAGGTCTCGTGGATGACGGGC
30 TGCTGATGCAGCAGACTCCGGCGCGGTCCGACGGTGATGGGACCTAAG
GTCCAGGGGGCCTTGCACTTGACACGCTGACACGCGAAGCGCCTCTTTC
CTTCTTCGTGCTGTACGCTTCTGCAGCTGGGCTTTTCGGCTCGCCAGGCC
20 AGGGCAACTATGCCGCGAGCCAACGCGTTCCTCGACGCCCTTTCGCATCAC
35 CGAAGGGCGCAGGGCCTGCCGGCGCTGAGCATCGACTGGGGCATGTTAC
GGAGGTGGGGATGGCCGTTGCGCAAGAAAACCGTGGCGCGCGGCAGATCT
CTCGCGGGATGCGGGGCATCACCCCGATGAGGGTCTGTGAGCTCTGGCG
40 CGCTTGCTCGAGGGTGATCGCGTGACAGCGGGGTGATACCGATCACTCC
25 GCGGCAGTGGGTGGAGTTCTACCCGGCAACAGCGGCCTCACGGAGGTTGT
CGCGGTGGTGACCACGACGCGCGCGGTGCTGATCGGACCGCCGGGGAT
CGGGACCTGCTCGAACAGCTTGCGTTCGGCTGAGCCGAGCGCGGGCGGG
45 GCTGCTGCAGGACGTCGTGCGCGTGAGGTCTCGCATGTGCTGCGTCTCC
CTGAAGACAAGATCGAGGTGGATGCCCCGCTCTCGAGCATGGGCATGGAC
30 TCGCTGATGAGCCTGGAGCTGCGCAACCGCATCGAGGCTGCGCTGGGCGT

5

10

15

20

25

30

35

40

45

50

55

CGCCGCGCCTGCAGCCTTGGGGTGGACGTACCCAACGGTAGCAGCGATAA
CGCGCTGGCTGCTCGACGACGCCCTCGTCGTCCGGCTTGGCGGCGGGTCC
GACACGGACGAATCGACGGCGAGCGCCGGTTCGTTTCGTCCACGTCCTCCG
CTTTCGTCTGTCTGTCAGCCGCGGGCTCGTCTCTTCTGTTTCACGGTT
CTGGCGGCTCGCCCGAGGGCTTCCGTTCCCTGGTCCGAGAAGTCTGAGTGG
AGCGATCTGGAATCGTGGCCATGTGGCACGATCGCAGCCTCGCCTCCGA
GGACGCGCCTGGTAAGAAGTACGTCCAAGAGGCGGCCTCGCTGATTACAGC
ACTATGCAGACGCACCGTTTGCCTTAGTAGGGTTCAGCCTGGGTGTCCGG
TTCGTCAITGGGGACAGCCGTGGAGCTCGCCAGTCGTTCCGGCGCACCGGC
TCCGCTGGCCGTCTTACGTTGGGCGGCAGCTTGATCTCTTCTTACAGAGA
TCACCCCGGAGATGGAGACCGATATAATAGCCAAGCTCTTCTTCCGAAAT
GCCGCGGGTTTCGTGCGATCCACCCAACAAGTCCAGGCCGATGCTCGCGC
AGACAAAGTCATCACAGACACCATGGTGGCTCCGGCCCCCGGGACTCGA
AGGAGCCGCGTGAAGATCGCGGTCCCTATCGTCGCCATCGCCGGCTCG
GACGATGTGATCGTGCCTCCGAGCGACGTTTCCAGGATCTACAATCTCGCAC
CACGGAGCGCTTCTATATGCATCTCCTTCCCGGAGATCACGAATTTCTCG
TCGATCGAGGGCGCGAGATCATGCACATCGTCGACTCGCATCTCAATCCG
CTGCTCGCCGCGAGGACGACGTCGTCAGGCCCCGCGTTCGAGGCAAAATG
ATGGCAGCCTCCCTCGGGCGCGGAGATGGTTGGGAGCAGCGTGGGCGCT
GGCGGCCGGCGGCAGGCCGCGGAGGCGCATGAGCCTTCTTGGACGTTTGC
AGTATAGGAGATTTTATGACACAGGAGCAAGCGAATCAGAGTGAGACCAA
GCCTGCTTTTCGACTTCAAGCCGTTTCGCGCCTGGGTACGCGGAGGACCCGT
TCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCCATCTTCTACTGGGAT
GAAGGCCGCTCCTGGGTCTTACCCGATACACGACGTGTCGGCGGTGTT
CCGCGACGAACGCTTCGCGGTCACTCGAGAAGAGTGGGAATCGAGCGCGG
AGTACTCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTG
TTCGGGCTGCCGCCGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCC
GTCGTTTACGTACGCGCCATCGACCTGCTGCGCGCCGAAATACAGCGCA
CCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAAGAGGAGTTCGACGTT
GTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCTCTGTT

5

10

15

20

25

15

30

35

40

25

45

30

50

55

GAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCCGA
CTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACC
AAGACCCTGCTCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGT
CCTCGATGAGCGGCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGC
TGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAAGGAGCTGGTC
GCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATAACCACGATCTACCT
TATCGCGTTTCGCTGTGCTCAACCTGCTGCGGTGCCCCGAGGCGCTCGAGC
TGGTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTC
CGCTTCGACAATATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGA
CCTGGAGTACTGCGGGGCATCGATCAAGAAAGGGAGATGGTCTTTCTCC
TGATCCCCGAGCGCCCTGAGAGATGGGACTGTATTCTCCAGGCCAGACGTG
TTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGAGGCCC
CCATGTCTGCCCCGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCG
TGGGCACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAACTCCC
GTGTTTGGATAACCACCCCGGTTCCGGAACATCGAATCACTCAACGTCAT
CTTGAAGCCCTCCAAAGCTGGATAGCTCGCGGGGGTATCGCTTCCCGAAC
CTCATTCCTCATGATACAGCTCGCGCGCGGGTGCTGTCTGCCGCGGGTG
CGATTGATCCAGCGGACAAGCCCATTTGTCAGCGCGCGAAGATCGAATCC
ACGGCCCCGAGAAGAGCCCGTCCGGGTGACGTCGGAAGAAGTGCCGGGCG
CCGCCCTGGGAGCGCAAAGCTCGCTCGTTCGCGCTCAGCACGCCGCTCGT
CATGTCCGGCCCTGCACCCGCGCGAGGAGCCGCCCGCCCTGATGCACGG
CCTCACCGAGCGGCAGGTTCTGCTCTCGCTCGTCCCTCGCGCTCGTCC
TCTGACCGCGCGCCCTTCGGCGAGCTCGCGCGGCGGCTGCGCCAGCCC
GAGGTGCTCGGCGAGCTCTTCGGCGGCGTGGTGCTGGGCCCGTCCGTCGT
CGGCGCGCTCGCTCCTGGGTTCCATCGAGTCTCTTCCAGGATCCGGCGG
TCGGGGTCTGTCTCTCCGGCATCTCTGGATAGGCGCGCTCGTCTCTGCTG
CTCATGGCGGGTATCGAGGTCGATGTGAGCATCCTGCGCAAGGAGGCGCG
CCCCGGGGCGCTCTCGGCGCTCGGCGCGATCGCGCCCCCGCTGCGCACGC
CGGGGCCGCTGGTGCAGCGCATGCAGGGCGGTTACGTGGGATCTCGAC
GTCTCGCCGCGACGCTCTGCGCAAGCCTGAGCCTCGGCGCCTGCTCGTAC

5

176

10

5

15

20

10

25

15

30

35

40

25

45

30

50

55

ACCTCGCCGGTGCTCGCTCCGCCCGCGGACATCCGGCCGCCCGCCCGGGC
CCAGCTCGAGCCGGACTCGCCGGATGACGAGGCCGACGAGGCCGACGAGG
CGCTCCGCCCGTTCCGCGACGCGATCGCCGCGTACTCGGAGGCCGTTTCGG
TGGGCGGAGGCGGCGCAGCGCCGCGGCTGGAGAGCCTCGTGCGGCTCGC
GATCGTGCGGCTGGGCAAGGCGCTCGACAAGGTCCCTTTTCGCGCACACGA
CGGCCGGCGTCTCCAGATCGCCGGCAGACTCCAGAACGATGCGGTCTGG
TTCGATGTCGCCGCCCGGTACGCGAGCTTCCGCGCGGCGACGGAGCACGC
GCTCCGCGACGCGGCGTCCGCCATGGAGGCGCTCGCGGCCGCCCGTACC
GCGGATCGAGCCGCGTGTCCGCTGCCGTAGGGGAGTTTCGGGGGGAGGCG
GCGCGCCTTACCCCGCGGACCGTGTACCCGCGTCCGACCAGCAGATCCT
GACCGCGCTGCGCGCAGCCGAGCGGGCGCTCATCGCGCTCTACACTGCGT
TCGCCCCGTGAGGAGTGAGCCTCTCTCGGGCGCAGCCGAGCGGCGGCGTGC
CGGTGGTTCCTCTTCGCAACCATGACCGGAGCCGCGCTCGGTCCGCGCA
GCGGCTAGCGCGCGTCCGCGCAGAGATCGCTGGAGCGACAGGCGACGACC
CGCCCCGAGGGTGTCGAACGGATTGCCGCGAGCCCTCATTGCGGATCCCCTC
CAGACACTCGTTTCAGCTGCTTGGCGTCGATGCCGCCCTGGGCACTCGCCGA
AGGTCAGTCTCGTCGCGCCACTCGGATCGGATCTTGTTCGAGCACGCGTCC
TTGCTCGAATACTCCCGGTCTTGTCCGATGTTGTTGCACCGCGCCTCGCG
GTCGCACCGCGCCGCCACGATGCTATCGACGGCGCTGCCGACTGGCACCG
GCGCCTCGCCCTGCGCGCCACCCGGGGTTTGGCGCTCCCGCCTGACCGC
TTTTTCGCCGCGCGACGCCGCGAGCAGGCTCATTCCCGACACCGAGATCAG
GCCCACGACCAGCTTCCAGCAATCTTTTGCATGGCTTCCCCTCCCTCAC
GACACGTCACATCAGAGACTCTCCGCTCGGCTCGTCGGTTCGACAGCCGG
CGACGGCCACGAGCAGAACCGTCCCCGACCAGAACAGCCGCATGCGGGTT
TCTCGAACATGCCCCGACATCCTTGGCGACTAGCGTGCCCTCCGCTCGTGC
CGAGATCGGCTGTCTGTGCGACGGCAATATCCTGCGATCGGCCGGGCAG
GAGGTACCGACACGGGCGCCGGGCGGGAGGTGCCGCCACGGGCTCGAAAT
GTGCTGCGGCAGGCGCCTCCATGCCCGCAGCCGGGAACGCGGCGCCCGGC
CAGCCTCGGGGTGACGCCGCAAACGGGAGATGCTCCCGGAGAGGCGCCGG
GCACAGCCGAGCGCCGTACCACCGTGCGCACTCGTGAGCTCCAGCTCCT

5

10

15

20

25

30

35

40

45

50

55

CGGCATAGAAGAGACCGTCACTCCCGGTCCGTGTAGGCGATCGTGCTGAT
CAGCGCGTTCTCCGCCTGACGCGAGTCGAGCCGGGTATGCTGCACGACAA
TGGGAACGTCCGATTTCGATCACGCTGGCATAAGTCCGTATCGCGCGGGATC
GGCTCGGGTTCGGTCAGATCGTTGAACCGGACGTGCCGGGTGCGCCTCGC
TGGGACGGTCACCCGGTACGGCCCCGGCGGGGTGCGGGTTCGTGAAGTAGA
CGGTGATGGCGACCTGCGCGTCCCAGTCCGACGCATTCAACAGGCAGGCC
GTCTCATGGCTCGTCATCTGCGGCTCGGGTCCGTTGCTCCGGCCTGGGAT
GTAGCCCTCTGCGATTGCCAGCGCGTCCGCCCCGATCGGCTTCTCCATAT
GTCCTCCCTGCTGGCTCCTCTTTGGCTGCCTCCCTCTGCTGTCCAGGAGC
GACGGCCTCTTCTCCCGACGCGCTCGGGGATCCATGGCTGAGGATCCTCG
CCGAGCGCTCCTTGCCGACCGGCGCGCCGAGCGCCGACGGGCTTTGAAAG
CACGCGACCGGACACGTGATGCCGGCGCGACGAGGCCGCCCCGCGTCTGA
TCCCGATCGTGACATCGCGACGTCCGCCGGCGCCTCTGCAGGCCGGCCTG
AGCGTTGCGCGGTTCATGGTCGTCTCGCGTCAACGCCACCCGCCGATTCA
CATCCACCGCGGCACGACGCTTGCTCAAACCGGGCGAGACGGCCGGGC
GGCTGTGGTACCGGCCAGCCCCGACGCGAGGCCCGAGAGGGACAGTGGGT
CCGCCGTGAAGCAGTGAGGCGATCGAGGTGGCAGATGAAACACGTTGACA
CGGGCCGACGAGTCGGCCCGCGGATAGGGCTCACGCTCGGTCTCCTCGCG
AGCATGGCGCTCGCCGGCTGTGGCGGCCCGAGCGAGAAAATCGTGACGGG
CACGCGGCTCGCGCCCCGGCGCCGATGCGCACGTGCGCCCGACGTGACC
CCGACGCCGCGACACGCGGCTGGCGGTGGACGTGTTACCTCTCGCCG
CCCAGCGCATCGAGGCCGGCAGCGAGCGGTTCTGTCGTCTGGCAGCGTCC
GAGCTCCGAGTCCCCGTGGCAACGGGTTCGGAGTGCTCGACTACAACGCTG
CCAGCCGAAGAGGCAAGCTGGCCGAGACGACCGTGCCGCATGCCAACTTC
GAGCTGCTCATCACCGTCGAGAAGCAGAGCAGCCCTCAGTCTCCATCTTC
TGCCGCCGTATCGGGCCGACGTCCGTCCGGTAACATCGCGCTATCAGCA
GCGCTGAGCCCGCCAGCAGGCCCCAGAGCCCTGCCTCGATCGCCTTCTCC
ATCATATCATCCCTGCGTACTCTCCAGCGACGGCCGCGTCAAGCAACC
GCCGTGCCGGCGCGGCTCTACGTGCGCGACAGGAGAGCGTCTTGGCGCGG
CCTGCGCATCGCTGGAAGGATCGGCGGAGCATGGAGAAAGAATCGAGGAT

5

178

10

5

15

20

10

25

15

30

35

40

25

45

50

55

CGCGATCTACGGCGCCATCGCAGCCAACGTGGCGATCGCGGCGGTCAAGT
TCATCGCCGCCGCCGTGACCGGCAGCTCGGCGATGCTCTCGAGGGCGTG
CACTCCCTCGTCGATACTGCAGACGGGCTCCTCCTCTGCTCGGCAAGCA
CCGGAGCGCACGCCCCCGACGCCGAGCATCCGTTCCGCCACGGCAAGG
AGCTCTATTTCTGGACGCTGATCGTCCGCATCATGATCTTCGCCGCGGGC
GGCGGCGTCTCGATCTACGAAGGGATCTTGACCTCTTGACCCGCGCCA
GATCGAGGATCCGACGTGGAACCTACGTCCTCTCGGCGCAGCGCCGTCT
TCGAGGGGACGTCGCTCATCATCTCGATCCACGAGTTCAAGAAGAAGGAC
GGACAGGGCTACCTCGCGGCGATGCGGTCCAGCAAGGACCCGACGACGTT
CAGCATCGTCCTGGAGGACTCCGCGGCGCTCGCCGGGCTCACCATCGCCT
TCCTCGGCGTCTGGCTCGGGCACCGCCTGGGAAACCCCTACCTCGACGGC
GCGGCGTCGATCGGCATCGGCCTCGTGCTCGCCGCGGTTCGCGGTCTTCCT
CGCCAGCCAGAGCCGTGGGCTCCTCGTGGGGAGAGCGCGGACAGGGAGC
TCCTCGCCGCGATCCGCGCGCTCGCCAGCGCAGATCCTGGCGTGTGCGCG
GTGGGGCGGCCCTGACGATGCACTTCGGTCCGCACGAAGTCTGGTCGT
GCTGCGCATCGAGTTCGACGCCGCGCTCACGGCGTCCGGGGTTCGCGGAGG
CGAGGGAGCGCATCGAGACCCGATACGGAGCGAGCGACCCGACGTGAAG
CACATCTACGTCGAGGCCAGGTGCTCCACCAGCGCGGAGGGCGTGACG
CGCCGTGGAGAGACCGCGCGCGGCTCCGCCATCCTCCGCGGCGCCCGG
CTCAGGTGGCCCTCGCAGCAGGGCGCGCCTGGCGGGCAAACCGTGCAGAC
GTCGTCTTCGACCGGAGGTACGCTGGTTGCAAGTCGTACGCGGTATCG
CGAGGTCCGGCAGCGCCGAGCCCCGGGCGGGCCGGGCGCACGAAGGCGCG
GCGAGCGCAGGCTTCGAGGGGGCGACGTCATGAGGAAGGCCAGGGCGCA
TGGGGCGATGCTCGGCGGGCGAGATGACGGCTGGCGTCGCGGCTCCCCG
GCGCCGGCGCGCTTCGCGCCGCGCTCCAGCGCGGTGCTCGCGCGATCTC
GCCCCGGCGCGGCTCATCGCTCCGTGTCCCTCGCCGGCGGCGCCAGCAT
GGCGGTCTGCTCGCTGTTCAGCTCGGGATCATCGAGCGCTGCCGATC
CTCCGCTTCAGGGTTTCGATTTCGSCCAAGGTGACGAGCTCCGATATCGCG
TTCGGGCTCACGATGCCGGACGCGCGCTCGCGCTCACCAGCTTCGCGTC
CAACCTCGCGCTGGCTGGCTGGGGAGGCGCCGAGCGCGCCAGGAACACCC

5

10

15

20

25

30

35

40

45

50

55

CCTGGATCCCCGTGCGCGTGGCGGCCAAGGCGGCCGTCGAGGCGGCCGTG
TCCGGATGGCTCCTCGTCCAGATGCGACGGCGGGAGAGGGCCTGGTGCGC
CTACTGCCTGGTCGCCATGGCGGCCAACATGGCCGTGTTGCGCTCTCGC
TCCCGGAAGGGTGGGCGGCGCTGGGGAAGGCGCGAGCGCGCTCGTGACAG
5 GACGGGCGCGGGCAGCCCCGCCATCGGAGGCCGGCGTGACCCGCTCCG
TCACGCCCCAGCCCGCGCCGCGTGATCTCCCGCGGACAGGGCGCGTACCG
TGGACCCCGCACGCGCCGCGTCGACGGACATCCCCGGCGACCCGCGCGGC
GCGACCCGCGCAACTCCGGCCCGCCCGGGCATCGACATCTCCCGTGAG
CAAGGGCACTCCGCTCCTGCCCGCTCCGCGAACGATGGCTGCGCTGTTT
10 CCACCCTGGAGCAACTCCGTTTACCGCGTGGCGCTCGTCGGGCTCGTCGC
CTCGGCGGGCGGCGCCATCCTCGCGCTCATGATCTACGTCCGCACGCCGT
GGAAGCGATACCAGTTCGAGCCCGTCGATCAGCCGGTGCAGTTCGATCAC
CGCCATCACGTGCAGGACGACGCCATCGATTGCGTCTACTGCCACACCAC
25 GGTGACCCGCTCGCCACGGCGGGGATGCCGCCGACGCCACGTGCATGG
GGTGCCACAGCCAGATCTGGAATCAGAGCGTCATGCTCGAGCCCGTGCGG
CGGAGCTGGTTCTCCGGCCACGCCGATCCCGTGGAACCGGGTGAACTCC
GTGCCCCGACTTCGTCTATTCAACCACGCGATCCACGTGAACAAGGGCGT
GGGCTGGCGTGAAGCTGCCACGGGCGCGTGACGAGATGGCGGCCGTCTA
CAAGGTGGCGCCGATGACGATGGGCTGGTGCTGGAGTGCCATCGCCTGC
20 CGGAGCCGCACCTCCGCCCCGCTCTCCGCGATCACCGACATGCGCTGGGAC
CCGGGGGAGCGGAGGGATGAGCTCGGGGCGCAGCTCGCGAAGGAATACGG
GGTCCGGCGGCTCACGCACTGCACAGCGTGCCATCGATGAACGATGAACA
GGGATCTCCTTGAAAGACGCAGATGAGATGAAGGAATGGTGGCTAGAAG
40 CGCTCGGGCCGGCGGAGAGCGCGCGTCTACAGGCTGCTGGCGCCGCTC
25 ATCGAGAGCCCGAGCTCCGCGCGCTCGCCGCGGGCGAACCGCCCCGGGG
CGTGACGAGCCGGCGGGCGTCAAGCCCGCGCGCTGCTCAAGCTGCTCG
GCGCGAGCATGGCGCTCGCCGGCGTCGCGGGCTGCACCCCGCATGAGCCC
45 GAGAAGATCCTGCCGTACAACGAGACCCCGCCGCGCTCGTGCCGGGTCT
CTCCCAATCCTACGCGACGAGCATGGTGCTCGACGGGTATGCCATGGGCC
30 TCCTCGCCAAGAGCTACGCGGGGCGGCCATCAAGATCGAGGGCAACCCC

5

180

10

15

20

25

10

15

30

35

20

40

25

45

50

55

GCGCACCCGGCGAGCCTCGGCGCGACCGGCGTCCACGAGCAGGCCTCGAT
CCTCTCGCTGTACGACCCGTACCGCGCGCGCGCGCCGACGCGCGGCGGCC
AGGTGCGCTCGTGGGAGGCGCTCTCCGCGCGCTTCGGCGGCGACCGCGAG
GACGGCGGCGCTGGCCTCCGCTTCGTCTCCAGCCCACGAGCTCGCCCCCT
CATCGCCGCGCTGATCGAGCGGTCCGGCGCAGGTTCCCCGGCGCGCGGT
TCACCTTCTGGTCGCCGCTCCACGCCGAGCACGCGCTCGAAGGCGCGCGG
GCGGCGCTCGGCCTCAGGCTCTTGCTCAGCTCGACGTCGACCAGGCCGA
GGTGATCCTCTCGCTGGACGCGGACTTCCTCGCGGACATGCCGTTACGCG
TGCGCTATGCGCGCGACTTCGCCGCGCGCGCGCGCCCCCGGAGCCCGGCG
GCGGCCATGAGCCGCCTCTACGTGCGGAGGCGATGTTACGCCCACGGG
GACGCTCGCCGACCACCGCTCCGCGTGCGGCCCCCGGAGGTGCGCGCGG
TCGCGGCCGGCGTCGCGGTGGGAACCTCGTGACAGAGTCTTTGTCTTGCGC
CCTGTCCGGGAATAACGGACACCTTATCGCGGGTCGCTCTTTGTGCGCGG
CTTCTGTACCTCTCAGGACAGGTAGAAGAGGGAAGTACAGGGGCCCTTATGT
TAAGTGGGGATGCCTTCGGGACGGCCGCAAATATATCCTATCACCTCACT
GGGTGTGGGGGAGCACCGCGAGGATGTACAACCTCTGTAAGTCTATGTGA
GATAATGTGTGACGTGATCTGAGACTTATTTGTGTGACCGAGACGTCTCT
CTTATTGGTACGCATAGTATAATATAACACGTCTCATACATACTCCCGAC
ATATCCGCGGTATGCGCGCACATAGAATAGGTGATGATAAATCCCTAGTG
TGTGGAAGTAGAAGATGCGGGAGTTACCTGATATTTACGGAAAAAGTATT
ATCTCAACTACCTCTCTGTTGAGACTATCACTTCGGTGTCGTTGTGCTGC
TGGT,

or its complementary strand,

(b) DNA-sequences which hybridise under stringent conditions
to regions of DNA-sequences according to (a) encoding proteins
or to fragments of said DNA-sequences,

5

10

(c) DNA-sequences which hybridise to the DNA-sequences according to (a) and (b) because of a degeneration of the genetic code,

15

(d) allele variations and mutants resulting by substitution, insertion or deletion of nucleotides or inversion of nucleotide segments of DNA-sequences according to (a) to (c), wherein the variations and mutants offer isofunctional expression products.

20

11. Peptide encoded by a DNA sequence according to claim 10 selected from the group consisting of

25

Seq ID No 83

>Contig56_003 2890 amino acids MW=307428 D pI=5.76 numambig=13

15 IRPRAAAVPMRSTVTGGVIAGPELGASYWADNLRQPVRFAAAAQALLEGGPALFIEMSPH
PILVPPPLDEIQTAAEQGGAAVGSLRRGQDERATLLEALGTLWASGYPVSWARLFPAGGRR
30 VPLPTYPWQHERCWEIEVPDARRLAAADPTKDFYRTDWPEVPRAAPKSETAHGSWLLLA
DRGGVGEAVAAALSTRGLSCTVLHASADASTVAEQVSEASRRNDWQGVLYLWGLDAVVD
AGASADEVSEATRRATAPVLGLVRFLSAAPHPFRFWVTRGACTVGGEPEASLCQAALWG
20 LARVAALEHPAAWGGGLVDLDPQKSPTEIEPLVAELLSFDAEDQLAFRSGRRHAARLVAAP
35 PEGDVAPISLSAEGSYLVTGGLGGLGLLVARWLVERGARHLVLSRHGLPERQASGGEQP
PEARARIAAVEGLEAQGARVTVAAVDVAEADPMTALLAAIEPPLRGVVHAAGVFPVRHLA
ETDEALLESVLRPKVAGSWLLHRLRLDRPLDLFVLFSSGAAVWGGKGQAYAAANAFDLC
40 LAHHRRAHSLPALSLAWGLWAEGGMVDAAHARLSDIGVLPMTGPAALSALERLVNTSAV
25 QRSVTRMDWARFAPVYAARGRRNLLSALVAEDERAASPPVPTANRIWRGLSVAESRSALY
ELVRGIVARVLGFSDFGALDVGRGFAEQGLDSLMALEIRNRLQRELGERLSATLAFDHPT
VERLVAHLLTDVLKLEDRSDTRHIRSVAADDDIAIVGAACRFPGGDEGLEYWRHLAEGM
45 VVSTEVPADRWRAADWDYDPDPEVFGRTYVAKGAFLRDVRSLDAAFFAISPREAMSLDPQQ
RLLEVSWEAIERAGQDPMALRESATGVFVGMIGSEHAERVQGLDDDAALLYGTTGNLLS
30 VAAGRLSFFLGLHGPTMTVDTACSSSLVALHLACQSLRLGECQALAGGSSVLLSPRSFV

50

55

5

182

10

15

20

25

30

35

40

45

50

55

AASRMRLSPDGRCKTFSAAAADGFARAEGCAVVVLKRLRDAQDRDPILAVVRSTAINHD
GPSSGLTVPSPGAQQALLRQALAQAGVAPAEVDFVECHGTGTALGDPPIEVQALGAVYGRG
RPAERPLWLGAVKANLGHLEAAAGLAGVLKVLLEHEQIPAPPELDELNPHIPWAEPLV
AVVRRAPWPARGARPRRAGVSFAFLSGTNAHVVLEEAPAVEPVAAAAPERAAELFVLSAKS
5 AAALDAQAARLRDHLEKHVELGLGDVAFSLATTRSAMEHRLAVAASSREALRGALSAAAQ
GHTPPGAVRGRASGGSAPKVVFVFPQGGSQWVGMRKLMAEPPVFRAALEGCDRAIEAEA
15 GWSLLGELSADAAASQLGRIDVVQPVLFAMEVALSALWRSWGVEPEAVVGHSMEVAAA
VAGALSLEDAVAIICRRSRLRLRISGQGEMALVELSLEEAEALRGHEGRLSVAVSNSPR
STVLAGEPAALSEVLAALTAKGVFWRVQKVDVASHSPQVDPLREELIAALGAIRPRAAAV
20 PMRSTVTGGVIAGPELGASYWADNLRQPVRFAAAAQALLEGGPALFIEMSPHPILVPPLD
EIQTAAEQGGAAGVSLRRGQDERATLLEALGTLWASGYPSWARLFPAGGRRVPLPTYPW
QHERYWIEDSVHGSKPSLRLRQLRNGATDHPLLGAPLLVSARPGAHLWEQALSDELSYL
25 SEHRVHGEAVLPSAAYVEMALAAGVDLYGTATLVLEQLALERALAVPSEGGRIVQVASE
EGPGRASFQVSSREEAGRSWVRHATGHVCSGQSSAVGALKEAPWEIQRRCPVLSSEALY
15 PLLNEHALDYGPCFQGEQVWLGTGEVLGRVRLPGDMASSSGAYRIHPALLDACFQVLT
LLTTPESIEIRRLTDLHEPDLPRSRAPVNQAVSDTWLWDAALDGGRRQSASVPVDLVG
SFHAKWEVMERLAQAYIIGTLRIWNVFCAGGERHTIDELLVRLQISVVYRKVIKRMEHL
VAIGILVGDEHFVSSQPLPEPDLA AVLBEAGRVFADLPVLFWCKFAGERLADVLTKGT
LALAILFPGGSFDMAERIYRDSPIARYSNGIVRGVVEAARVAPSGMFSILEIGACTGA
20 TTAAVLPVLLPDRTEYHFTDVSPIFLARAEQFRDYPFLKYGILDVDQEPAGQGYAHQRF
DVIVAANVIHATRDIRATAKRLLSLLAPGGLLVLVEGTGHPWFDITGLIEGWQKYEDD
LRIDHPLLPARTWCDVLRVGFADAVSLPGDGSPAGILGQHVLRSRAPGIAGAACDSSGE
SATESPAARAVRQEWADGSADVHRMALERMYFHRRPGRQVWVHGRLRTGGGAFTKALAG
40 DLLLFEEDTGQVVAEVQGLRLPQLEASAFAPRDPREEWLYALEWQRKDPIPEAPAAASSSS
25 AGAWLVLMDDGGTGAALVSLLEGRGEACVRVIAGTAYACLAPGLYQVDPAPQPDGFHTLLR
DAFGEDRICRAVVMWSLDATEAGERATAESLQADQLLGSLSALSIVQALVRRRWRNMPR
LWLLTRAVHAVGAEDAAASVAQAPVWGLGRTLALHPELRLCTLVDPNPAPSPEDAAALAV
45 ELGASDREDQVALRSDGRYVARLVRSSFSGKPATDCGIRADGSYVITDGMGRVGLSVAQW
MVMQGARHVVLVDRGGASEASRDALRSMAEAGAEVQIVEADVARRDDVARLLSKI EPSMP
30 PLRGIVYVDGTFQGDSSMLELDAARRFKEWMPKVLGAWNHLALTRDRSLDFFVLYSSGTS

5

183

10

LLGLPGQGSRAAGDAFLDAIAHHRCKVGLTAMSIWGLLSEASSPATPNDGGARLEYRGM
EGLTLEQGAALGRLLARPRAQVGVMLNLRQWLEXXXXXXXXXXXXXWYNLLII IQYTK
VPPQGPFRL*

5 Seq ID No 84

15

>Contig56_027 700 amino acids MW=80569 D pI=7.02 numambig=0
MNMELNITSKSNPFGDTTAENDKKMLSNAFIETADFRTLIETDDRTIVVGRRTGKSALF
IQLNEHWKKDKKILILSFSPDDSQIIGFRSMLKPFTGSFNLARAATRLWRYAMLMEIAS
YISSHYKLSSQISSETLLNEHLKKWNSAQGDILRKCLVAKEYLDENNPEESIGDLQFNL
10 NISEIENNIVSLLERSDRKVILMDKLDEAYEPDNIGIGIIAGLAYASIELNQKAKCIRP
IIFLRDNIFRSLSKEDPDYSRNIQGVIRLHWDWAQLLMLSAKRMKVAFKLDIEKDQVRW
DRCTADDLKGNGFKRCLQFTLYRPRDLLSLLNEAFFSAFRENRETIINTDLEYAAKSIS
25 MARLEDLWKEYQKIFPSIQVITSAFRSIEPELTVYTCLKKIEASFELIEENGDPKITSEI
QLLKASGILQSLYSVGFVGIRDKNTSSYSFCHDGRTPDKGFESNEKLLIHPCYWLGLNLN
15 RNALAPEEAEEINDEYDINIISDNSAIRNKTIGQITTHLDQIPIGNEGATEFEQWCLDAL
RIVFASHLTDIKSHPNGNAVQRRDIIGTNGGKSDFWKRVL EDYKTRQVVFDKKNFEELGP
30 SEYRQLQSYLTGPYKGLGFIINRDESEVLKSGKDL DWTKEYQSHNSLI IKLPKYISKL
LQKLRNPEKHDAIDRQMGKLLTLYETSYMAIKSTQKKRRK*

25

30

20 Seq ID No 85

35

>Contig57_001 372 amino acids MW=38411 D pI=12.39 numambig=10
MLTSXXXXXXXXXXLLAYRCATAARGAGIRHHRQYAQRRRTTVVLYARAAGTLPDRRHG
LLVVARGHPPCLPQPARSRERSRAGGRRQHAPFVQDDDNAGAHPGAVARWPLPDIRRLGQ
40 RVRPWGGLRYGRAQTALRRPATRRSDLGSDPGFGHESGWFPVDRVDGTQCARSGGALARGA
25 AERSRRRRGHRLCRDPRNGDLARRPDRGRGAACRVGAGAGRWEPLRAGRSEDKPRPPGGR
CRRGGFDQGGAGSAPRTDPAKPPFPHAQSADPDRGDRARAGDGAGAVAAGGPTALRGGER
VRFQRHQRPCRAGGGAGHGARTGDAGALSGAFGAVGEERRRAGRTGGAALSAHRRVPGAG
45 SRRRRVQPGIDA*

50

55

5

184

Seq ID No 85

10

>Contig57_002 2259 amino acids MW=238258 D pI=5.92 numambig=0

5

MSYTLGLQGPCLTVDTVCSSSLVAIHLACRSLRARESDLALAGGVNMLLSSKTMIMLGRI
QALSPDGHCRTFDASANGFVRGEGCGMVVLKRLSDAQRHGDRIWALIRGSAMNQDGRSTG

15

LMAPNVLAQEALLREALQSARVDAGAIGYVETHGTGTSLGDPiEVEALRAVLGPARADGS
RCVLGAVKTNLGHLEGAAGVAGLIKAALALHHELI PRNLHFHTLNPRIRIEGTALALATE

20

PVPWPRAGRPRFAGVSFGLSGTINVHVLEEAPATVLAPATPGRSAELLVLSAKSAAALD
AQAARLSAHIAAYPEQGLGDVAFSLVSTRSPMEHRLAVAATSREALRSALVAAQGGTPA

25

GAARGRAASSPGKLAFLFAGQGAQVPGMGRGLWEAWPAFRETDFRCVTLFDRELHQPLCE
10 VMWAEPCSSRSSLLDQTAFTQPALFALEYALAALFRSWGVEPELVAGHSLGELVAACVAG

25

VFSLEDAVRLVVARGRLMQALPAGGAMVSTAAPEADVAAAVAPHAALVSTAAVNGPEQVV
IAGAEKFVQQIAAFAARGARTKPLHVSHAFHSPMLDPMLEAFRRVTESVTYRRPSIALV

30

SNLSGKPCCTDEVSAPGYWVRHAREAVRFADGVKALHAAGAGLFVEVGPKPTLLGLVPACL
PDARPVLLPASRAGRDEAASALEALGGFWVVGGSVTWSGVFPSSGRRVPLPTYWPQQRERY

35

15 WTEAPVDREADGTGRARAGGHPLLGEVFSVSTHAGLRLWETTLDRKRLPWLGEHRAQGEV
VFPGAGYLEMALSSGAEILGDGPIQVTDVVLIIETLTFAGDTAVPVQVVTTEERPGRRLRFQ

40

VASREPGERRAPFRIHARGVLRRIGRVETPARSNLAALRARLHAAVPAAAIIYGALAEMCL
QYGPALRGLAELWRGEALGRVRLPEAAGSATAYQLHPVLLDACVQMIVGAFADRDEAT

45

PWAPVEVGSVRLFQSRSPGELWCHARVVSDDGQQASSRWSADFELMDGTGAVVAEISRLVVE
20 RLASGVRRRDADDWFLELDWEPAALGGPKITAGRWLLLGEGGGLGRSLCSALKAAAGHVVV

50

HAAGDDTSTAGMRALLANAFDGGQAPTAVVHLSSLDGGGQLGPGLGAQGALDAPRSPDVDA
DALESALMRGCDSVLSLVQALVGMDLRNAPRLWLLTRGAQAAAAGDVSVVQAPLLGLGRT

55

IALEHAELRCISVDLDPAEPEGEADALLAELLADDAEEVALRGGDRLVARLVHRLPDAQ
RREKVEPAGDRPFRLEIDEPGALDQLVLRATGRRAPGPGEVEISVEAAGLDSIDIQLALG

60

25 VAPNDLPGEIEIPLVLGSECAGRI VAVGEGVNGLVVGQPVIALAAGVFATHVTTSATLVL
PRPLGLSATEAAAMPLAYLTAWYALDKVAHLQAGERVLIHAEAGGVGLCAVRWAQRVGAE

65

VYATADTPENRAYLES LGVRYVSDSRSGRFVTDVHAWTDGEGVDVVLDSLSGERIDKSLM
VLRACGRLVKLGRRDDCADTQPGLPPLLRNFSFSQVDLRGMMLDQPARIRALLDELFGLV

70

AAGAISPLGSGLRVGGSLTPPPVETFPISRAAEAFRRMAQQQHLGKLVLTLDDEVRIRA
30 PAESSVAVRADGTYLVTGGLGGLGLRVAGWLAERGAGQLVLVGRSGAASAEQRAAVALLE

75

80

5

185

10

AHGARVTVAKADVADRSQIERVLREVTASGMPLRGVVHAAGLVDDGLLMQQTTPARFRTVM
GPKVQCALHLHTLTREAPLSFFVLYASAAGLFGSPGQGNAAAFLDALSHHRAQGLP
ALSIDWGMFTEVGMAVAQENRGARQISRGMRGITPDEGLSALARLLEGDRVQTGVIPITP
RQWVEFYPTATAASRRLSRLVTTQRAVADRTAGDRDLLEQLASAEPSARAGLLQDVVRVQV
5 SHVLRLPEDKIEVDAPLSSMGMDSLMSLELRNRIEAAALGVAAPAALGWYPTVAAITRWL
15 LDDALVVRLLGGGSDTDESTASAGSFVHVLFRFPVVKPRARLFCFHGSGGSPGFRSWSEK
SEWSDLEIVAMWHDRLASEDAPGKKYVQEAASLIQHYADAPFALVGFSLGVRFVMGTAV
ELASRSGAPAPLAVFTLGGSLISSSEITPEMETDI IAKLFFRNAAGFVRSTQQVQADARA
DKVITDTMVAPAPGDSKEPPVKIAVPIVAIAGSDDVIVPPSDVQDLQSRITTERFYMHLLP
20 GDHEFLVDRGREIMHIVDShLNPLLAARTTSSGPAFEAK*

20

10

Seq ID No 87

25

>Contig57_027 419 amino acids MW=46737 D pI=5.09 numambig=0
MTQEQAQSETKPAFDKPFAPGYAEDPFPALERLREATPIFYWDEGRSWVLTRYHDVSA
15 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID
LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRFRFGSATA
30 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS
TKELVALVGAI IAAAGTDTTIYLI AFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI
LRIGTVRFARQDLEYCGASIKKGEMVFLIPALRDGTVF SRPDVFDVRRDTGASLAYGR
20 GPHVCPGVSLARLEAEIAGVTIFRRFPKMLKETPVFGYHPAFRNIESLNVILKPSKAG*

35

Seq ID No 88

40

>Contig57_043 492 amino acids MW=52617 D pI=11.54 numambig=0
MAARARKSCRAGSRPAPMRTSPPTSTPTPRPRGWRWTSFTSRRPSASRPAASGSSSGSV
25 RAPS PRNGSECSTTTLPAEEASWPRRPCRMP TSSCSSPSRSRAALSLHLLPSSGRRPS
GNIALSAALSPPAGPRALPRSPSPSYHPCVLLQRRPRRSNRRAGAALRARQESVLARPAH
RWKDRRSMEKESRIAIYGAIAANVAIAAVKFIAAAVTGSSAMLSGVHSLVDTADGLLLL
45 LGKHRSARPPDAEHPFGHGKELYFWTLIVAIMIFAAGGGVSIYEGILHLLHPRQIEDPTW
NYVVLGAAAVFEGTSLIISIHEFKKKGQGYLAAMRSSKDPTTFTIVLEDSAAAGLTIA
30 FLGVWLGHRLGNPYLDGAASIGIGLVLA AVAVFLASQSRGLLVGESADRELLAAIRALAS

50

55

5

1.86

10

ADPGVSAVGRPLTMHFGPHEVLVVLRIEFDAALTASGVAEERERIETRIRSERPDVKHIY
VEARSLHQRARA*

12. DNA sequence according to any of claims 1 to 5 wherein the
DNA is selected from the group consisting of

15

(a) the following DNA sequences:

Seq ID No 89 (>Contig10)

20

10 GGTAGTGAAATATGCTGTATTCAACAGAAAGCTTGATGAATTGATCTAGA
AAGTAGAGCGAGAGAATCAAGTAAGATAGTAGGATGCATTATAAATATAG
AATATATACTGCATACGATGACAGCATGCGCACGAATAGAATGCATAAGA
25 GGCAAGCCAATAACCAAAAGTGGAGCCAGAGGAGATAGTCTCGCCAGTAG
AAATAATGCTCAGCCAAGCGAGGTTGGACATATCAGTTCCAGAGTAGGTC
15 TCAACCCCGTATATGAGTCCAATGAAGCCTGTCTCATCCAGTTAACGGCC
TTTGAGCAGAGAATCCTCCCTATTTTCGGAGAGGACGCGTCGAATATAA
30 AGCAGGTCCAAAGAAGCAAGCAATAGCCAAAAGTTTGAAAGGTTAGTACG
AGCAGCGGCTGGAGGACACTATGGTCGTGCAACGGGGGTAAAGGTTTCA
CGTATTGTAGCAGAGCACGTCAGAGGGTTATTCGTGACATTCGAGGCCAA
20 CGAGGCGGTAGGACTTCGTAAGCGCATGACCATCCCGGTCACAAACGTAG
35 TCGGAGCGCCTCGTCAAGCTCAACAAGGCCCTAGAACGCGCGGCGCAGA
TCGACCCTTTTAAACGCCGGCACCAGCCGACCGTCCTGCCAGGTTGT
AAAGCGCTCCATCGGCCGACTTATGGCACTCGAGCCAAATCGCCCGGTTTC
40 CCCATCGGTGAGCGCAAACGGCCCCCGGGCGTCGCCACCCGCGCGGAC
25 GAGGGGCGGTCCAGACGGGTGATCTCTCTCGTGAGCTCGCGGAGAGAGCC
TCCTCGCAAGATCGATGTCAGCGGGATCGCGCGCCCGTCCGCACCTGAA
ACCGGTGCTGGAGCTCGACGGCAGCGAGGGAGTCGAGGCCGAACCGCGAT
45 ATCGGCAGCGCGTCGTGATCTGCCCCGGCGTCCAGACGAAGCGCGCGGGC
GAGGGTCGAGCGCAGCGGTCCAGCAGGCTCCGGCCGAGGGCTCCTCGG
30 TCTCCGGGGGCGCGTCGTCCGGGGCGAGGCGTCGTGAGGAGCTCCGGC

50

55

5

187

10

5

15

20

10

25

15

30

35

40

25

45

30

50

55

GCGAACGCGACGTGGCGCTCGCCGAGCGCGTCCTCGAGAAAGGCGCGCCG
GCACTCCCTCCGGCGGACCTTCCCGCTCGACGTCTTCGGCAGCGCGCCCG
GCGCGATCAGCGCGACGGCGTGCGCGACGAGCTGGTGCTCGGCGGTACCC
GCCTCGCGCACGGCCGCCACGATCTCGCGCGGATCCGCGGCCACGCGCGG
GTGACCTCGCACACCACGGCGAGGCGCTCCTCGCCCTCGTGCTCCACGG
AGAACGCGGCGCTGCAGCCCGGCCGACGGCGCGATGGCTGCTCTCGACG
GTCTTCTCGATGTCTCGGGGAAGTGGTTGCGGCCTCGAAGGATGATGAG
GTCTTTCGACCTCCCCACCACGAACAGCTCGCCGCCCCGAGGAAGCCGA
GATCTCCCGTGCGCAGGTAGCGCGGCGCCGCGCTGCCAGCGAGCGTGCC
CCGAACGTGGCCTCCGTCTCTCCGGGCGCCCCAGTAGCCGACGGCTAC
GCTGGGCGCGGACACCCAGATCTCCCGATCTCCCCCGGCCGAGCTCGT
TCCCCGCGGGATCGACGATCGCGACCGCCCGGGATCGAGCGCCCGACCG
CTGCCGACGAACACGCGCGCGCCCTCCGCGCCGACGCGACGGCGCGCCC
GAGTCCACCTCCTCGGGGCGAGGCGCGCCAGCACCGGCGCCTCGGCCC
GCGCTCCGCGCGCTCACGATGAGCGTGGCCTCGGCGAGCCCGTAGCAGGGA
TAGAACGCCTCTCGCCGGAACCCGCTGACCGCGAAGGCGCGCGGAAGCG
ATCGAGCGTGTGGCGCGCACCGGCTC3GCGCCCGTGAACGCGACCTCCC
ACGACCGCAGATCGAGCGCCGCTCGCTCCTCCTCCGAGCTCTTCCGGACG
CACAGGTGCTATGCGAAGTTGCGGCGCGCGCTCACCGAGGCGCCGAGCGC
CGACACGGCGCGGAGCCACCGCATCGGCCTCTGCAGGAACGAGAGCGGCG
ACATGAGCGCGACCGGATCCGCGCGTAGAGCGCCTGCAAGATCCCGCCG
ATGAGCCCCATGTCGTGATACGGCGGCAGCCAGATCACCCGACCGGATC
CGGGCTCGTCAGGTGCAATCCATGCGCGATGAGCCGCGAGTTGTGCAGCA
GATTCCTCGTGGGTGAGCATCACCCCTTGGGCTCGCCGGTCGAGCCGGAG
GTGTATTGAAGGAACGCGACCGACTCCGGCCGAGCGCCGCGCCCGGCC
CTCGATCGGGCCCGCGACGGGCGTGGTTCGCGATCCACCGGAGCCGCT
GCAGCGCGGCGCCGCGCGCTGGCCGCGAGGACGCCACGATGCCGGCG
ACGGCCGATGACGTGAGCGCCCGCTCGGCGCGCGCTCCGCGACGATGGA
AGCGACGCGCGGCAGCGTCCGCTCGAGCCGGCCGAGATCCGGCGGATAGG
CGGGCACGGTCCGGAATCCAGCGTAAAGACACCCGAAGAACGCGGTGATG

5

188

10

15

20

25

30

35

40

45

50

55

TACTCGATCCCCGGCGGATACAGCAGCAGCGCGCGGGCCCCGGGGGCGAC
GCCCCGATGCCTGCAAGAGGGCCGCGACGGTTCGCGCGCGCTCGTCAATTT
CCCCGAGGGTCACCCAGGTCGCCCCGGCCTCGACGTCGCCGGA CTCAAGA
AAGCAATAGATTGGCGGGCGGGCTCAGCTTCGGCCCCGCTGGCGCAAGAG
5 GTCGATAACGGTGAAGGGCGGTTCCGTTCCGTTCCAATGCAAGAA
AAGCATCATTGTAACAGACCCCTCCGCCGCGGAGATAGCAGCTTGTC
CGCTGCGACACAACCGCCGCGGACGCGCGTGGCACGGCGGGATCCGGGC
GTACTCCACCTGCACTTCCCGTCGCGTCACGCTCGCTCCGCCGCGGGTG
TCGTGAACCACCGCCACAGCGACACGCGCCCTGGCTCCGAGGGCGGCGT
10 CCACGTGGTCCCGTTGCCCCCGGGAGGCGCCGGTCGTATCGCTACCA
GGCGGCGCGCCCCGTGAACTCGCCACCGTCCGTGTAATAGTCGACCCAG
ATCGCCTCGCGCGCCGGTGGACCGCGAGCCCGCGGGCTTCCTCGTCCAC
CTCGGCGGCTTCTCGGGCACGACAGCCTCAATCTCATAGGTCGTGCACT
25 CGTCTCGGCCGCTCGGTCCGGCCGCGAGCCTTGGGCTGCTCCTCGGAC
CGAACGCACCGCTTCACGACGGGCAAACCGTCCTCGCCCCGGCGCGACCTC
ATTGCCATCGAGCTTCAGCGTGAAGCCGTCGATGGGCGGGTTCGTGTTCA
30 GCCGCTCCTTCTTGAAGACATAGACCTGCGTGTAGCCACGACGAAGCTG
TCCGGACCGAGCACCGTCCCGTCGTGCGCGACGCACTCCAGCGGAAACCC
GGCCGTTTCGGGCGCCGAAGCCACGCGTGTGTCGTCGGCGCGCACACGGCGA
20 ACAGCACGTAAGCCGACGAGTACACCGTCCCCGTCTCGGTGGGCCTCGCG
TCCTTGAGGATCTCCTTGGGCAGCTTCACCCGAACGAGACCGCATCGGG
CTCGCCGCTCTTCTCCGGACCGATCTCCTGCTGCGCGAAGGGGACGGTGC
GCTCCCCGTGCGCGCTGCCGCTGCCGCCGTGCGCGCTGCTGCCGCCGTG
40 CCGTGGCACCGCCACCGCCCGCCCGCCAGCGCCACCATTGCCGCCTTC
25 GCCGCTCCAGCGCCACCACTGCCGCCGTGCGCGCGCTGCCACCATCGC
CACCCTGCCGCCCGCCGCTTCCGCCGTGCCGGCCGGCACCGCCTCC
CGGATTCGCGACGATTCCCACCGCATAGGTGCCAGCCACTGCGGGATGC
45 ACCCGAGGTGCTCGTCCACCCGACCGGCGGATTACGCAGCCGCCACC
CACGTGACCTCGACCTTCCGCGGCGCGCCGCTCCGCGCCTTTCGCGTC
30 GGCGTACGTCATCCGGAACGTCACGAGCTCTCCGCCCGCGGTACGGCT

5

10

15

20

25

30

35

40

45

50

55

TGTCCGCCGTCACGGCGAGGACGCGGAGCCCCCTTCACCTCGGACGAAGGG
GCCATGTCGCTCCCGGCGCAGCAAGGGATGCCCACGGCCAGGGTCGAGAG
CGCGGCCAGCAGCGCGCGACGAGCGGGCAGTGCGGTCCGTTTCATCAGAA
AICTCCTCGCAGCCCGAGCGTGCGGCAGGAAGGGGAGCCCCGTCACGTACT
5 CCCGCTTCGTGTAGTTGAAGTTGTAGCTGATGCCCTCCGAGCCATGTAA
TTGTAGACGTTCTGGATATCGAGGTAGAGCCCAGCTGCCACCTCTTGAA
TTTCCACGTCTTGTGCGCGCGGATGTCGAGCTGGTGAACAGCGGCATCC
GCTCGCTGTAGTCACCCCCGAGCGGGATCGGCGAATACCTCGCCGAGGAC
GCGTGGTAGATCGCGTTCACCCGGTTCGGATTGCACCCCTTCTCCTCCGG
10 ATCGCAGACATAGGGCGTCTGCAGGTTGCCCGACACGAGCCGGAAGCGCG
CGCCACGTCTCCAGCCCCGGCCGAGCCGAGGCTCCCGAGCACCCGTCAGC
ACGTGCGTCTGATCGAACTGGGTGAGGTGCTCCTCCTCGTCGGGGCCGTC
CTTGCGCACCGACCGCGAGAGGGTGTACGCCGCCAGCCGAAGAAGCGCT
25 CGTCCGGCTTGTACTTCAACAAGAGCTCGCCGCCGACCGCGTATCCGGTG
CCATCGTTGGCATAGTCGTCTTCTCCGGCGAGAAGACGACCAGCCGATC
GAGCTGCTTGTAGAACCCGTCCAGCGTCACCTCGATCTGCGGCGTGATCT
30 CCTGCTCCACGCCGAGGCCGTAATGCACGGCGCGGTTGACTTGAGCTCC
GCATTGCCGAACGGCTCGATGCTCTCCGCGAACTGCGGCGCCTGATAATA
AAGGCCACGCCCCCTTGGCCGTCTCGCGGGAAGCCGCTCCGATGT
20 CGTAGCCGCGGTTGACCCGCGGGCTCACGTGAGCGTCTGCGTATCGAGC
GCGTAGTCGACCCGACCCCGGGGACGATCCGCGCCCGGGCGAGGGGAC
GACCTCGAGCTCGGCATACGCCGCGGGCCGCGAGTACGCGCCGTCGAACG
ACCGATCCTGGAACGGGTACGTCGAGAACGGCTGGTTCGACGGGTGGCCC
40 GCGGGCTGCTGCGACGCGCGCGGATGTTGACCGTGGCGACGCCGCCGA
25 GAGGTGCGTGCCGACGTTTCATCGTGAGGTACCGCGGAACCTGTGCGAGA
GCTCCAGCCGCGAGGTCGAGCGAGGTCGAGACGACGTTGAAGGCGAGGGGA
GAGATCTCGAAGTCGGCGATGTCCCGGCCGAGCGCCATCGACCACAGCAG
45 CCGATCCCGGCTCCCGATCCGGTTCTCGTAGCTGAGCTGGAAGCGCTGGA
AGGCGGTGTGCAGCCCGAAATCGCCCGTCAGCGCCGCTCGTCCTCCGGC
30 GGCTTGTCCAGGGTGATCTTGAAGGCGTCGTCCGATCCGTAGAAGCTCGC

5

190

10

5

15

20

10

25

15

30

35

20

40

25

45

30

50

55

GCCACGCGCTCGCTCGCGGAGGGGCGGCCCTCGAGGACGAACTGGTAAT
CATAGTAGACGGGCGCCTGCGTGACGCTGGAGCCCGCCTCCTTGAGCACG
GGCCCGAGCCACGCGTCGACCCAGCTGCGGCGGCCCGCCGCGATGAACGT
CCAGTCCTTGAGGAACGGGACGGGGCCCTCGAGGAGCACGCGCCCGTCGA
TGAGGTCGAGCTGGACCACGCCGTGGTACTTGCCGTCCTGCTTCGGCGAG
CGGAGCCCGACGTCGACGATGCCGCCCATGGCGCGGCCGTACACGGCGCT
GAAGTTGCCCGGATAGAAGTCGATCTTCTCGAGCATCTCGGTCGGCACGA
CCGAGGAGAGGCCGCCGAAGTGGTAGATGATCCGCACCGGGGTGCGATCG
ACGAACGTGAGCGTGTCCTGGGGCGCGGACCCGCGCACGATGAGCAGCCC
GAAGCCGCTGCGCGGACGCCCGGAGGCTCTGCAGCGACCCAGCGCGT
CGCCCGCGGTGCCGGGATGCGGTGATCTCGCGGCGCTCGATCGTCCTC
CGCGTCACCTCGCGCGGCGGGCGCTCGCCCTGCACGGTCACCTCGATGCC
CGGCGCCTTGCCGTCCTGCGGCGCGGCGAGCGAGATGCGGTAGCGCACCT
CGATCGCCTCGCCGGCCGCGATCTCCTCCTCGGCGGCGAACGGCTCGAAC
CCCGCGGCGGCGACCTCGACCGGTAATTGCCGGGGGGGAGATTCTTGAA
GCGGAATTGCCGCCCTGGTCCGTCTTCGCCTCCTCGCGGCCCGCTCGG
GGCGCACGAGGGTGACCGCATGTCCGGAGCGGCTCGCCGGTGCCCGCG
GACAGGACGGTCCCGACCACCGTCTCGACGTGCGCGGGCGCGCCGGCGC
GGCCGCATCGGCGGGCTTGGGCGTGAGCGTGAACGCGTACCGGTAGAGGA
TGCGCGCCCGCGGGCGTGCCGTCCGGGCGCCGCGCCGGCGCGAACTCC
AGGCCGGGCGCGGCCCGGAGCGCCGCTCGTTGAAGCCGTGCCCGCCGG
GCGTCGCGACCTCGGCCCTGGTGACGCGCCCGGTCTTGTGATGTGAGC
TTGAGGATGACGCTGCCCTCGACGCCGGCGCGCTGGGCCCTCGATCGGATA
CGCGGGCGGGAGTACTTGATCAGCGTCGGCGGGCTGATGGCGGCGGGCG
CCGGCGGGGCGCGCCCGGCTGAGGGACGACGACCGCGCCGGCGCGCCG
CGGGGACCGAGGCGCCCGGAGTCGCCCTCGGCGGCGAGGAGGCTC
GGGCGGGGCGCGCCGGCGGGCTGCGCGCGCGCTGCGCTGCCGGTCATCG
CGACCGGAGCAGCAGCGCTTCCGAGACGACGAGCGCATCACGAGGAC
GCTGTGGAAGGCATGCGGCCCGCCCTCTCGCATGGCGAGGCCGAGGCGGA
AAGACGCATCGCGCAGCCAGGACCCTGCTTACATTGCTTCACACAACGG

5

10

15

20

25

30

35

40

45

50

55

GCGCCGCGCGCTCCCGGGCCGCGCAGCGCAGGCGGCGCGCGCCCCG
CGGGCGGCGCGATCGCGAGCGGCGCGCGGTGCGATCAGCCGCCGACCTCG
GCCACGAACCGGCTCACGTCGTCGCTGTCGCCCACGAGCAGCAGCGTGTC
GCCGTCGCGGATCACGTAGTCCGGTGTGGGCGCCTCGAGCCGCGGCTTGT
5 CGCCGGGCCGCTTGTTCGTGTGCGGCCGCACACCGAGCACGTTGATGCGG
TACCGCTGGCGGATCTTCGAGCCGGCCAGCGTCTGCCCCACCAGCGGCCC
GTGGGCGTTCCAGGGGACCACGCGGTAGTGCTCGCGAGGTCGAGGAGGT
CCTGCGCGAGCGGCATGGTGATGTGCGCGCCGACGCGGCGGCCATCTCG
GTCTCGAGCTGGATGACGCGGGTCGCGCCCCACCGCGCGCAGGATGTGCGC
10 CTGGCGATCGGTGGCGGCGCGCGCGATGATCTCGCGCACGCCCATCCGGA
CGAGGGAGGCCACGCAGAGCACGACGGCTCGAAGTGCTCGCCGAAGGTC
ACGATCGCGGTCTCCACGTACTGCGCGCCGATCCCCTCGAGCACCTTGTG
GACGGTGGCGTCGCCGACGAACGCGGCCGAGGTCTTGTCTTCACGGCGT
25 CGACGGCCTCCGGTTGTTGTGACCGCGATCACCTCGGCCCGGTTCTTC
15 CAGAGGGTCTCGACGACCGACGTGCCGAACCGCCCCGAGCCCCGATGACGA
GGACGCTCTTCGATTTCATGGTCTCCGGTCGCGCGCCGCCCTCGGGGC
GCGGCGCCGGAAGGTCTCACGGATGCGCCGGGAGCGCCACGGTTCGGCG
30 CTGCCGTCGTCCGACGCGCGCGGCCCGCGGGCCGCGCGCGCGCGCT
CAGTAGAGCTCGTCTGGTGCTGCCACCGCTCCGAGATCCACGCCTTCAG
20 GTACGCGATCTCCTCCTCGTACGTCTGTAAGTCGTGCGCCAGCTCCAGC
CCTCGTAGCTCCGGTACGCCTCGCCCCACCGCGCCTCGTCCCGGCGCGCG
35 CTCGCGTCGATGCGCTCCACGTAGCCGTCCACGATCGCGTGGATCTCGGC
CTCGCGGAGCGCGCCGCGCAGGACCTGATCGTAGCGGGCGCGCAGCGGGT
CGCCGATCGACGGCTCCTCGAGGAGGCGCTCGAAGAGGAGGTTACGTCG
40 CGGTAGTCGACGCGATCCGACGCCGGCTCGCGCTCGGTCTCCACGACTG
25 GCCGAAGCTCGCGTTGAAGTCCCACGGCGCGTAGCGGAATACGCCGTCCG
CGGCCGGATCGCGGTAGTGGTAGCTGTTCTTTCCGGCCGAGTCGTTGGCC
45 ACGATGAACGTGACGAAGATCCACCAGTCCTCGTAGTCGCGCAGATCGAT
CCGCGACCCGATCTCGGCGGCGAACGTGGCGTCGTGCGACTCGGCCACGA
30 AGCTCACGAGATCTTCCAGATCCGAGAACGCCTCCGGCTCGCCCTCGGCC

5

192

10

GGCGCCCTTCCTTCTCTCGAAGCCGTCGTGCAGCGTGTCTTGGGGTC
GCCGGACCGGTCCGTACGCGGAAGTTCGCGTCGTGGCTGACCGCCTTGT
AGAGGTTGCCGTCTTCCGGGTAGCCGTGGTCTCCATCAGGTAGCCGTCTG
ACGTGATCCGCGACGGTGTAGAGCCCCGCTACTCCCCGTGAGGTACAG
5 GACGGCGCTGTAGGTCTTGATCTGGATGTGCTCGGGATCGAGGCGGTCC
AGAGGTCATAGGCGAGGCGCTGCCGGACATAGGAGTTGTCGTGCAACGTC
15 GTGATGAGCACGACCTTGCGGCGATCGGTGAAGCCGCCCTCGTCGGG
CTCGTTGAACTTGTCTCTTGGGGAACTTGAGGGTGTAGCTCCGCTTCG
GGTACGAGAGCGAGCTCTCGCCGCGGAGCTCCGCCTCCGCGGCGTACGTG
20 TGGCCGCGGTAGATCACCCTGGCCGGGGCGTACTCCTTGTCCTCGGGGAC
10 GGGCGAGAGGAAGAGCACCGGCAGGCCGTACTCCTCGGGGTAGCGGGTCG
GATCGACGACGGGCACGTTTCGACGGATCGGCGAAGGCGTCGGCGACGCCG
ACCTTGACGCGCCCCGACCTCGGACGTCTGCGCGACCGGATCTCGATGTC
25 GTAGACGGCGGCTGATCGAGCCCCGGGCGAGAACGTACCTCGCGCGGA
15 TCGGGTCTGACGCGGCGCCCTCGGGGAGCGGGCCGACCTCGAACGCGTCG
CCGGCGAGCGCGAGGCCGCTCGCGCACGTACCGGGAACGTACCGGTCTC
30 CCCCTCGAGGAGCCAGTSCGGGCCCGCCCCGACGGCTGSCAGCGCGAGC
CCTCGGCGCTGGAGCCCCCGCGCTGGAGCCGGAC

25

30

35

20 Seq ID No 90 (>Contig11)

40

GGCGACCCACATATCACATAGTAGAATCAGTGTGAGTTAGACAAATGTC
GAGTGATGAGAAGGACAGAAGTGACAACTCTGTCTGATCACTGTAGAACGA
GAGAGTATGAGCCTGCATACATGATAGCGGACATGAGAACGAGTGTANTA
TGATGCTACTAAGAGAGTAACAGATCAGAGACTAGAGTAGAGCAATAGAA
25 NTCAGAGATAAGTCAATGACGAGGAGTAGTGATAGAGCTCTTAATAATGG
CTGAGGTCTGAAGATAGAAGTGATAGAGCGATAGATATACAATCGGTTGA
AGCAGAGAGTAAGATAAGATCAGACACNGAGTACAGAGAGAGACGAATAG
45 ATGGCGTGATNTCACAGAGAGGTGCGAGCGTAGCTGACGAGAGCAGAGAC
GCAGAGTAAGTCACACCTAGATAGTTACGGCGAGAGACAAATGATAGGAA
30 GGAGTGACGAGATCAACAGNCCGGAGCACAGAAGCGTGAGATGCGACCG

50

55

5

10

15

20

25

30

35

40

45

50

55

TGTAATAAACAGGAGACAAGAGCGACTACATAAGAGAGCGAAGCGAATAG
ATAAGATATAAGCCCAGAGCAAAATAGAAGGAGAGAGAGAGTATTTGTAA
TAAAGCAACAAGACGGAGAGAGCGAAGCAGCAGGCAACGATTAGAAGAAA
GACGACAGGAAAGTGAAAGCGAAAGAGAGCAGGTAGAAAGAGAACCAAAA
5 AAGCACGAAGGAAAAGGAAGCTTCTATGATAGGTGCGGGACAAGGCGTAG
CTACAGGAGACAGCCGGCATAACGAGGAGCCGGTAAAAGCTAGCCTTTCAG
AACACATCGGGAGCGCGTAAAGGCGGACCACGCTCGACGGGATCATGTAC
GCCGACAGCGACGCCTTCAGCCCCGCGCGCACGTCCGGCGCGTCGCCCCG
CGCGTCGCCGTGAGCAGCAGTAGGCGACCAGGCGGGCGTCGCCCCGGC
10 CGTCTCGCGCAGGACCACGGCCGCCTGGCCCCACGCCGGGCACGCGCCG
ATCTCGCCTCGACGTGCGCGAGCTCGATCCGGTGCCCCCGA3CTTGAT
CTGGTGGTCCGAGCGGCCCTGGAACTCGAGCATCCCGTCGGGCAAGAAGC
GCGCGACGTGCGCGGTCCGGTACATCCGCCCGCCGCGCGCGCGCGCAC
25 GGGTCGGGCAGGAAGCGCTCCGCGGTGAGCCCGGGCTGCCCCACGTAGCC
GCGCGCGAGCGGCGCGCCCCGCGATGTAAAGATCGCCGAGCGCGCCGATGG
CGGGGCGGCGCATCGCGCCGTGAGCAGCAACACCTCGGCGTTCGCGACC
GGCGCGCCGAGGGGGACCCACGTGACCCGCGGGTCGCTCGGCAGGACGCA
30 GCCGGTCACCGCGATCGCGGCCTCGCTCGGCCCGTACATGTTGATGAGGT
CGCCGTCTGTCTTCGCGTAGAAGCGCCGGACGAGATCGAGCGGCACCGCC
TCGCCGCCCCACGAGGACCTTCCGCGAGGCTCGCGGGGAACGGCTGCTCGGG
35 CCCCCGAGGAACGCCGCGAGCATCGAGGAGACGAAGTACGCGGTCTGTCG
CCCCCTCGTCGCGCACGAGGCGCCGAAGGTAATCGGGATCGCGGTGCCCCG
CCGCCCCGGGCGACGACGATCCGCGCGCCGAACGAGAGGGGCCAGAAGAT
CTCCAGACGGAGACGTGCAAGCCGAACGCGGCCTTGAGCAGGACCCGGT
40 CGTCCGCGGTGAGCGCCAGTACCGCTGGATCCACTGCATCTGGTTGACG
25 ATGGCGCGGTGGGAGATGAGGCTCCCTTCGGCGTGCCCGTCGATCCGGA
CGTGTAGATGACGTACGCGCCGCTGTCCGGCGGCGGGCTACGGCGGGCC
GCGCGTCGGAGCACGCGGCGATCTCGGCGGCCTCGCGTCGAGGAGCAGC
45 GTCGTCCAGCCCGCGGTGCGGAGCTCGTCGGCGATCGCGTCGTGCGTGAC
GAGGAGGCGCGCCCGCGCGTCCCGCATCATGAAGGCGAGGCGCTCGCCGG
30

5

194

10

5

15

20

10

25

15

30

35

40

25

45

30

50

55

GGTACTCGTGGTCGAGCGGCAGGTAGGCGCCGCCACCTTGAGCACCGCC
AGGGTCGCGACGACCATGTCTCGGAGCGCGGCACGCAGACCCCGACGAT
CGTGTCGAGCCCGACGCCGCGGCGCGCAGCAGCTCGCGAGCCGGTTCG
CGCGCCGCTCGAGCTCGCCGTACGTGAGCGACTTGCCCTCGCTGCGCACC
GCGACGACGTGCGGGTGCTGCTCGGCGCGCTCCTCGAACCACCGGTGCAG
CGCGCAGGCCGACGGCAGCTCCATCGCGGGGCCGCGCAGCACGCCTCGA
TCTCGGCGCGCTCGCCGGGCGGACGTACTCGCCCTGGGCGACCGGACGC
TCGGGTGGCGCGACAGGTCTCGAGCAGCGCCGCGAGGCGCTCGGCGAG
GCGCTCGGCGTTCGCGCGCGCGCCCGGCCGACGCGTTCGTAGCGGAGCTCCA
GCGACGCCGACGGGCCCGCGCCGCGCGCACTGCAGCCGGCAGGCGACCTGG
TCAGACGTGCTCCAGACGTTCGAGCACGCGGGCCCGCGCGCCGTTCAGCGA
CAGCGCCGGCGCGGCCTCCGCGGCGGAGAAGCCCCAGCTCATCCGGTGG
CTCACCCCGGGCGCGCGCTCTCTGGTTCGCGGCGCCCTCGGCCTCGGCGAG
CGCGAGCCGCCGCGCAGCTCGGCGAGCGTGTCCGAGGCCGAGATCTCGA
TCCGCACCGGCAGGAACCGCGGAACGGCCCCACCGCGCCCGCGAGCGCG
TCCAGCGACCGCCCGTCGAAGCGGACGGCCACGGTGACCTCGGGGCTCGTT
GCCCGCGCTCATCCGCCACAGGAGCGACGCCACAGGGCCAGGAGCACGA
TCCGCTGCGGGACCTGCCACGACGACGACCAGCGCTCGACCTGCGCCATC
CCGCCTTGTCCAGATCGACCCGCGCGCGCCCCGAGCCGGCGCCGGCGCC
GGCGCCCGCGCGGCTGAAGGCGAGGTGGAGCGGGGGCCGAAATGCGAGC
GGCGCTCGGCCCAGAACCTGCGCCCGTTCGCCGGCGTCTCCGACTCGAGC
ATCCCGTTGAGCCACTCGGCGACGTCCGCGTACTGCTGCTCGGGCGGCGC
GCCCCGCGCCCGGTCGACGCGCAGAGCTCGCGGACGAGCGGGGCGATCG
ACTCCTCGTCGACGCACACGCGGGCGCCGCGAGCACGAGCCGGCGCTCC
TCCGGGCCGACGCGGACCAGGCCGACGCGCAGCCCGTCTGTCGCGCCGCG
GTCTCCGAGAGGCGCGCGACGACCGCGACATCCGCTCGCCCTGCTCGG
CTTCGGAGCACCCGACCCAGTCGTCTGTGACGCCCACGCGAAGCGCGG
CTCGCCGACCACTGCGCGGCCTCGCCCGCCCGCTCGACGAGGCGCGT
GCGCAAGATCTCGTGCCGCTCGGCCAGCGCGAGCGCCGCGCCGAGAGCC
GTCCCTCGTCGCACGGGCGGTCAAGGCGACGACGCCAGCGTCCGGCAC

5

195

10

5

15

20

10

25

15

30

35

20

40

25

45

30

50

55

CCGGGCGCCCCGCTCCCGGTGAGCGCGCGGATCGCCCGCTGCTGCGG
CGAGAGGCTGAAGCCGGTCATCTCGTGGTCACTCATCCAGGTCGTCCTTC
GGTGAGGTCTTCGCTTCGCCCCGGCGCGCCCGGCGCGGAGGGTCACGGC
GCJCCGGCGCGCGGAGGTCAGCTTGTGAGCGCGGCGCCGCGCGCGCCT
TCCGCTCGAGCTCCCGGCGGGCCGCGCGGCGCGCTCGAGTCCCCCGG
AGCTCCGAGACCGGGGTGTCCGGCCGCGCCGTGCGAGTGGCCAGGATCTG
CCGTAATCGCTAAGGAAATTGTGACCGTCGCCGCCCGGTACAGCTCGC
TGCTGTGCTCGACGCCGAAGCGGAACGAGCCGCCGGCCTCGGCGACCGTG
AGGACGAAGTCGAACGCCGTGCTGGTTCGCTCGCCCTCCAGCGCCTCGAG
CTCGAGCCCTCGAGCTTCATCGGGGGACGTGCACGTTGCGCATGACGA
ACTTCGCGTCGAAGAGGGGCACGTGCCCCGACGGCCCCCTTCGGCCGAGG
GCCTCGACGAGCCGGTCGAACGGCAGGTCTGGTGTCTGAACGCCTCGAG
CGCGACGTCGCGCACGCGGCGGACCAGCGCGCCGAACGTGCGGTGCCCCC
CGCAGTCGGTCCGAGCACGAGCTGGTTGACGAAGAAGCCGATCATCGGC
TCGGTCTCGACCGGTTCCGGTTCGCGACGTCGGTGCCACGACGAGGTC
CTCGAGCCCGGTGCGCTGGTGCAGGACGAGCTTGTACGCGCGAGCAGGG
CCATGAAGGGGAGATCGCCTCCCGCTCGCAGAAGCCTTGATCTGGCGG
GTGAGCTCGGCCCCGGCTCGAGGCTCCGCCGCGCCCCGCGCCACGTCCT
TCGCCCCGCCGGCTCGTGGTCGACCGGCACGCGGGCCCGGCGCAGCGCGC
CCGAGAGCTTCGTGCTCCAGTACCAGGCTCGCCCTCCAGGACCTCGCCG
GACAGCCACGCCCGCTGGGCTGCGGCGAAGTCGACGTACTGCGCCGGGAG
CTCCGGCAGCCGGAAGGCTGGCCCTGCGCGAAGCCGCCGTAGAGCGCGG
CGAGCTCGCCGACGAAGACGCCGACCGACGAGCTCGAACACGACGTGG
TGCACGACGAGCGCGATGACGTGCTCGTCTGTCGCTTCCGATGACCCG
CACGCGGAGGAGCGGCCCGGCTCAGGTGGAACGGCGCGAGGCTCTCCT
CGAGGACGAGCGCCGAGACCGCCGCTCGAGGGCTCGCCCGGAGGTGC
TCGAGGTGCGACATCCGGAACGGCACCCGGGCTCGGGCGCGACGACCGG
GAACGGCACGCCGTCCCTGGCGCTGAACGTGTCGCGAGCGCCTCGTGGC
GCCGCGGATCTCGAACAGGCTGCGGCGGAGCGCGTCGACGTCGAGCCGG
CCCGTCGCGCGACACGAACGGCATGTTGTACGCCGGGCTGCCCGGCTC

5

10

15

20

25

30

35

40

45

50

55

GAGCTGATCGACGAACCACAGCCGGTGCTGCGCGAACGACAGCGGGAGCG
GGCCGTGCGGGGGATCCGCGCGATCGGGGGAACTCGCGCCGCGCGCC
TCGCGCGGGCGCGGGCGTCGACCTGGGCGCGAGCGCCGCGACCGTCGG
CCCCTGGAAGAGCGCCCGAGCGGGAGCTCGACGCCGAGCTGCGCGCGGA
TCCGGGACATCACCTGGGTGCGGACGAGCGAGTCGCCGTGCAGGCCGAAG
AAGTCGTCGTGGACGCCGATCTCGTGGACGCCGAGGAGGGCGCTCCCAGA
TCGCGGGCGATCGCGCGCTCGGACTCGGTGCGACGCGCGGGCGAACGCGGC
CCCGGCGTGCGCGGGAGACCGCCGACGTGCGGAGGGCGTTTCGGCGCGG
GCGCGAGGGGGCGCTCGGCCGCGCCGCTCGGCGGGGGGGCGCCGCGG
CGCGGCGCGATCCAGTGCCGCGCCCGCTCGAACGGGTACGTGCGCAAGCG
GACGAGCGCGCCGGGGACGCCCCGCGGACGGGCCGTCCAGTCGACGGC
GTGGCCCGCCTCCAGAGCTGGCCGAGGGCCTCGGCCAGGCTCGCGGGCT
CGGACGCGCGTGGGTGCGACCCGAGGCTCGCGATCGCGGCGCCGCCGCGC
CCGCCAGCGTCTGCCGCAACAGCGTGGTCAGCCCGCGGCCGGGGCCGAC
CTCGAGGAACAGGGCGTGCCCGGACGCGAAGAGCGCCTCGACGCCGTGCG
TGAAGCGGACCGGCTGGCGGAGGTGCCGCGCCAGTAGGCCGGATCGGTG
GCCTCGGCGTGGTGAGGAGGGCGCCGGTGACGTTGAGACCACGGGGAT
CTCCGGCGGGGAGAGCCGCGCGCGCCGACGCTCTCGAGGAACGGGGCCA
CCGCGCCGTGATGAGCGCGCAGTGGAACGCGTGGGACGTCTGCAGCGGC
CGGGCGAACACCTCGCGCGCCTCGAGGCGCGCGGCGAGATCGCGGATCGC
GCTCGCCGGGCCCCGACAACCGTGAGCTTCGGGCTGTTGACCGCGGCGA
TCTCCAGGCCGGCCTCGAGGAGGCCCTCGACGTCCGCGGCCGCGAGGCCG
ACGGCCAGCATGCTCCCGGCCGGCGCCGCTGCATGAAGCGCCCCGATC
GATGACCAGGGACATCGCGTCCTCGAGCGTGAACACGCCCGCGACGCAGG
CCGCCACGAGCTCGCCGAGGCTGTGGCCGATCATCGCCGCGGGCTCGATC
CCCCAGCTCATCCAGAGCCTGGCGAGCGCGAGCTCGACGGCGAAGAGCGC
GGGCTGCGCCAGCGCGGTGCCGAGCAGCGTGCGCCCGTCGCCCTCGCCCT
CGCGGAAGACGACCTCGCCGAGATCGAGCCGCGCGCCCGCGCCCGCCG
GCGCACGCGTTGAAGGCGCTCCGGAACGCCGCCTCCTGCGCGTAGAGCGC
GCGGGCCATCCCGACGGCCTGCGCGCCCTGGCCCGGAACGCGAAGACGG

5

10

15

20

25

30

35

40

45

50

55

GCGCGGCTCATCGGGGCGCGGAGCGCGCTCGCCCCCTCGCGGGCGAGCC
CCTGGATCGCCTCGGCGCGCGTCCGGGCGACGACCGCCCGGCGGTACGGG
TGCTCCGCGCGCCCGGTCTGGAGGTTGAACGCGACGTCGTCGAGCGGGAC
GTCGGTCGCCTCGAGGTGCGCGGCGAGCTGCGCGCAGGCCGTCGACAGCG
5 CCTCCGCGGTGCGCGCCGAGAGCGTCAGCACGTGATCGCGCTCCGGGGCC
GGGCGCGGGGCGGCGAGCGGGGCGGCTCCTCGAGCACGACGTGCGCGTT
CGTCCCGCGGATCCCGAACGAGCTCACGCCCCGCGGCGCGGGCGGAGCT
CGCGGGGCCAGGGCGCCGCTCCCGCGGACGAAGAACGGGCTCGCCGCG
AGGTCGAGCTTGGGGTTCGGCGCCTCGAAATGGACGCGAGGGCGGGATCTC
10 GCGGCTCCGCGACGACGTGCGCGCCTTGATGAGGCCCGCGACGCCCGCG
CGGCGTCGAGGTGGCCGATGTTTCGCCTTGATCGAGCCGAGCGCGCAGTAC
GCCTTCCTCGGGGTCTTGCGGCGGAAGGCCTGCGTGAGCGCCTCGACCTC
GATCGGATCGCCGATCGCGGTGCGGTCGCCGTGGGCCTCGACGTAGCCGA
25 TCGAGCCGGGATCGACGCCGCGGCGACCGACTGCGCCTCGGAGATCGCCGCC
GCCTGGCCGTCGACGCTGGGCGCCATGAAGCCGACCTTGCGCCCGCCGTC
GTTGTTGACGGCGGAGCCCCGGATCACCGCGTGGACCGTGTTTCGGTTCGC
30 GGAGGGCGTCCGCGAGGCGCTTCAGCGCGACGATGCCGACGCCGCTGCCG
CCCACGGTCCCTCGGCGCGCGCTCGAACGGCCGGCAGCGGCCGTCGGG
GGAGCAGATGCTGCCGGGCACGTACGGATAACCGCGCTTCTGCGGGATGC
20 CGATGGAGACGCCCGCCGCGAGCGGAGATCGCACTGGCCGCCGAGGAGG
CTCTCGCACGCCATGTGGACGGCCACCAGCGACGTGAGCACGCGGTCTG
CACGACCACGCTCGGCCCGTGGAGGTGAGTTTGTACGAGACCCGCGTCG
CGAGGTAATCCTTCTCGCTCGCCAGCATGAGCGCGTGCGGATCGACGGTG
40 GCCGCGAGATCCGGGTGCGAGAGGAGCTGGAGGAGGTACGTGTTGGAGCC
25 GCACCCCCGAAGACGCCGATCGCGCCCGGAACCGGGCCGGATCGCAGC
CGGCGTCCTCCAGGGCGGCGACCGCGCACTCCAGGAAGAGGCGTGCTGC
GGGTCCATGAGCTGCGCCTCGCGCGGCGAGTACCCGAAATAGGACGCGTC
45 GAAGCGGTGATGTCGTGAGCAGGCCGCCCGCGCAGACGACGCGCGCCC
CGGGGGCGCGCTCGCGCCGACCGCGGCTCCTCGCGCTCGCTC²CCGGG
30 AAGCGCGGATCGACTCGACGCCGCGCCGACGTTCTCCAGAGGGCGTC

5

198

10

5

15

20

10

25

15

30

40

25

45

30

50

55

GACGCTCGGGGCGCCGGGGAAGCGGCCCGCCATGCCGACGATCGCGATGT
CGTCCCCCGTCCTCGGTCTCGATCGGCTCTGACATGGCTATCCTCGCC
CCCCGGCGGCTCGCGCGTCGCGGCGCGCCTCGGCGCGCTGCGCCCCGACG
TCGGCCGGCTCGGCCTTGACCGTCGCCGCGTCGAGCCGCTGCGCCAGTTG
CTCGATGGTCGGGTACTGGAACAGGTGGTTCAGCGACACGGCCTGCGCCG
CGGCGCCCTCGTCGGGCGCGCGCGCCGCGATGCGCTCGGCGAGCAGGCGC
TGCGCGCGCACGAGGAGCAGCGAGGTGAAGCCGAGCTCGAAGAGGTTGTC
GGTCACGCCGACGGCCTCGACCTGCAAGACCTCCGCGAGCACCGAGGCGA
TGAGCCGCTCGGTTCGCGGTCCGCGGGGCGACGCGCGCGCGCGCGCGCG
ACCGCGGCGGGATCCGGCAGGGCGGCGCGGTCCACCTTGCCGTTTCGCGCT
CAGCGGCAGCGCCGGGAGGACGACGACCTCCGCGGGGATCATGTACTCCG
GCAGCTTCTCCGGACGAAGTCGCGGAGCGCGCGCCATCGCCGTCGGCG
CCGACGACGTACGCGACCAGGCGCTTCTCGCCCGACGGATCGGTCTTCGC
CGCCACGACCGCCTGCTCGACCGAGGGTGCTGCGCGAGGGCGGCCTCGA
TCTCGCCGAGCTCGATGCGGAAGCCGCGGATCTTCACCTGATGGTCGGTG
CGCCGAGCAGCTCGATGGTCCCCTCGGCGAAGTAGCGGCCAGGTCCGC
TGTCCTGTACAGCCGCTCGCCGGTCTGTTGGGTGCTTCAGGAACCGTCCC
GGGTCCGCGCCTCGTCGCGCCAGTATCCGAGCGCGACGCCGATCCCGCCG
ATGTGGATCTCGCCGGGGACCCCGATCGGACACGGCTCCAGCCCCCTCGTC
GAGCACGTAGGTGTGCTGGTTTCGCGAGCGGGCGGCCGTAGGGGATGCTGC
GCCACGCCGGGTCGACGTCCGCGATCGGCTGGGCGATCGACCAGATCGAC
GCCTCGGTTCGCGCGCCGAGGCTCACGACGCGGGGCGCGCGCAGGCCGC
GCGGATGCGATCGGGGAGCTTCAGCGGGATCCAGTCGCCGCTCATCATGA
CGAGGCGGAGCGACGACAGCGCCGGTTCGCCCGCGCCGGGGACGCGTCC
ATGAGCATCTCCATCAGCGCCGGGACCGAGTTCACACGGTCACCCGCTC
GCGCTCCACGAGCTCGCGCCAGTGCCCCGATCCGAGCGCGGGTACGGT
CGGGGATCACGACGGCGCCTCCGGCGGCGAGCGTCCCGAACACGTCGTAG
ACCGACAGGTCTGAAGCTCAGCGACGAGACGCGAGCACCCGGTCTCCGG
GCCGACGTCTGAAGCGGCGGTTGATGTCGAGGACCGTGTACCCGCGCCGC
GGTGGTCGATCATCACGCCCTTGCGCAGCCCCGTGGACCCGGACGTGTAG

5

10

15

20

25

30

35

40

45

50

55

ATCACGTAGGCCAGGTCGTCCGTGCTTCCGCCGGGCGGCCGGCGCGAC
GGGCTGCTCGCGCCACCGCTCGTCCGCGTCGACGGCGAGGCGCTCGATGC
CCGCGGGCCAGGCGATCGTCCCGTCGACCGCCGACTGCGTGAGGACGAGG
CGGACCTCGGCGTGCTCCAGGAGGTGCCTGAGGCGCTCCTCGGGGAGGCG
5 AGGGTCCAGGGGAGGTAGGCGGCGCGGCGCGCAGCACGCCGAGCACGG
CGGCCACCTGCTCCCAGCCCTTCTCCATGACCACGGCGACGAGCGCGTTC
GCGGTGCTCCGGAGCGCGAGGCCGCGCGCGGATCGCCTCGGCGCGCCG
GGCGAGCTCCCCGTAGGTGAGGCGCGCTCGGCGTCGACGACCGCGCACG
CGTCGGGTGCTCGACGGCGCGCTCAAAGAACGGCTCCTCCAGCCGGAGG
10 TGATCCGGGGTTGCGACCGCGGTGTCGTTCCACGCGACGAGGGCGCGCTC
GCGGTCTTCCGGCGCGACGGAGAGCGCGCGGACGCGCTGCGCGGGTCTCT
GAGTGGCGCGGAGAGCACGCTCTGCATCGTGGCGAGCATCCGGTCGATG
GTCCGCCGCTCGAAGAGGTGCGACGTTGTACTGGAGCGAGATCACGTCGCG
25 CCCGCCGCGGGCTCGACGCTGAAGCGCAGGTGGAAGCGGTGGCCTCGA
CCGGGAGATCGAGCGGCTCGATCCGCACCTCGCCGAGCTCGAGCGCCTCG
15 GTTGGGGCGTTCTGCACGACGAGCATGACCTGGAACAGCGGCGAGCGGCT
CAGGTGCGGGCGGGGTTGACCGCTCGACCACCTTCTCGAACGGGGCGT
CCTGGTGCTCGAACGCCTCGAGCGCGACCTTCCGCGCCCGGAGAGGAGC
TCCTCGAAGGTGCGGTGCGCGCCGAGGTGAGGCGCATGACGATCGTGTT
20 CACGAAGAAGCCGACGAGGGGCTCGAGCTCGGGGCGAGGCCGTTGGCGA
35 CCGCGGTCCCGATGGCGAGGTGCTCCTGGCCCCGAGCTGCGCCGAGGAGC
ACGCCGAGGGCGGCGAGCAGGACCATGAAGCGGTGGCGCCGCGGCTCCG
GGCGAGCTCGTCGAGCTGCGCCACGAGGCGCGCTCGAGCGGGAGGACCC
40 GCTCCGCGCCGCGGAACGTCTGGACGGGCGGCGCGGTGATCGGTCTGG
25 AGCTCCAGGACCGGCAGCCCGGAGGGTCCGTGTCCAGTGAGCGAGCTT
GTCCGGCAGCCGCTTCCCCGCGAGGTGGCGGCGCTGCCACACCGCGAAAT
CGACGTAAGGAGCGGCGAGCTCGGGCATGTCCGCGGGCCCGCGCCCCGC
45 GCGCGCCGGTAGAGCTCCGCGAGATCGCGGACGAGGGGTTGGAAGGACCA
GGCGTCCGTGACGATGTGGTGCCTGGACAGGACCAGGACGACGACGTCGT
30 GGTGAGGCGGAACAGCCTGGCGCGGAACAGGGCCCGCGCGGAGGTCC

5

200

10

15

20

25

30

35

40

45

50

55

AACCCCGTGGCCTGCTCGCGGACGCCCAGGCGCGCGCCGCGGCCTCCGC
CTCGTCCGGGGCGTGCCGCGGAGGTCCGACCACCTCCGCGGGGGCCGCT
CGGGCTCGCAGATCTTCTGCGCCGGCGTGGGGCTCGCGACGAACACCGTG
CGCAGGCTCCAGTGCCCGCCGACGAGCGCGGAGCGCGGAGGACAGCGC
5 GTCGACGTCGACGAGGTTCCGCGAGGCGGACCGCTGCACCACGTTGTAGG
CGGTCCCGCCGGGAGCAGCTGCTCGAGGACCCACAGGCGCTCTTGCTCG
TACGAGAGCGGATACGGCTCGTCCGCGGCGCGCGGCCAGCGAGGGCGC
GATCTCGCTGGCGGGCACCGTCGCTCGGCGGCGGTGGTCGAGGCGGCGC
CGGAGGAGAGGCGATCGGCGAGCTGGTGGAGGGTTGGGTGCTCGAAGAGC
20 GTGCGGAGGGTGGTGCGGATGCCGAGGAGGACTCGATGCGTCCGAGGAC
CTGCATGGCGAGCAGGGAGTGGCCGCGAGGTGGAAGAAGCTGTCGTGTC
GTCCGACGCGGTGAGGTGGAGGACGGATTGCCAGATGTGCGCGAGCTCC
CGCTCGAGCTCGCCCGAAGGGGGCTCGTAGTCGGCGTGCGCGGCGGTGG
25 CGCAGGGAGGAGCTTCTTGTGACCTTGCCCGAGAGGGACATGGGCAAGG
CGGGGAGCAGGACGAAGTGGGCGGGCACCAAGGCGTCGGGCACCAGGCGG
GCCATGCCCTCGCGCAGGTGCGGCTCGGAGGGCGGGTCGGCGCCCGGCAC
30 GACATAGGCAATCAGGCGCGCGGCGCTGCCTTGCCGTGGAGGACGACGA
CGCCCTCGCGGACGGCGGGCAAGCGTCGCGAGGGCGGATTGACCTCGCCG
AGCTCGACACGGCGACCGCGGAGCTTGACCTGCTCGTCGCGGCGTCCGGC
20 GAAGGCGAGCTGTCCGTGCGGGCGCCAGCGCACCAGGTCGCCGGTGCGGT
AGAGGCCTGCGCCGGGCTGGCCGAAGGATCGGGCAGGAAGCGCTCTGCG
GTCAGGTCCGTGCGTGTGTAGCCCTGGGCGAGGCACGCTCCGCCGATGTA
CAGCTCGCCGAGGACGCCGGGCGGGACGGGCTGCATGTGCGGGTCGAGGA
40 CGTAGACGAGGGCGCTGTCCGATGGGTGCGCCGAGCGGGGGCTCGTCGCCG
25 AGGTGCGCGACCTCGGCCGACGGTGGTGATGACGGTGGCCTCGGTGGGGCC
GTACATGTTGAAGAGGCGGAAAGGAGCGGTGCGCGGAGCGGATGGAGCT
TGTCGCCCGCGACGGTCATCGCGCGCAGGGCGATGCCGGTCCAGTCTTGC
45 TCGAAGCACGCCTCGGCCAGGGGCGTGCCCATGAACGAGAGCGTGGCCCG
CTGAGCGACAAGCCAGGAGACGAGCGCTGTGGGAGAGCGGAGCGCGTCGT
30 CGTCGGCGAGGAGGAGTGACGCGCCGAGGCGAGCGGCGTCCAGATCTCG

5

201

10

15

20

25

30

35

40

45

50

55

TAGACGGAGCGCTCGAAGCCGCTGGAGGCCAGCTGAGTCCAGCGATCGCG
GGGTGAGAGCGCGAGCAGGTGCTGGAAGAAGGAGACGAGCCTTGAAAGGC
TCGCATGGCGCACACAGACGCCCTTCGGCGTGCCGGAAGAGCCGGAGGTG
AAGAGGACATAGGCCAGGTCTGTCGGGCCCTGGAGACGAGAGGAA7GTGGGT
5 GCTGGGCGCGCACGCCCCGTCCTGGACGAGGTGGACGGGGCAGGGGGCGG
CGGTGAGCTTGTGGCTGGCCTGGCTGCTGGTGAGCACGAGCGGGCGCGG
CAGTCGGCGAGCATCTCGGCCAGGCGCGCCGGGGGTTGGCGGGTTCGAG
CGAGGCATAGGCGGCGCCTGCCTTGAGGACGGCGAGCTGGGCGGCGACCA
TGCGGGGCGAGCGCTCGATGCAGACGCCGACGCTGCCGGGGCCGACG
10 CCGCGGTGCGCGAGCCACAGGGCGAGCTCGCTGGACCAGGTGCTGAGCTC
TGCGTAGGTGAAGCGCTGGTGTCCGAACTCGAGCGCCGTGGCGTCCGGCT
GTCGAGCGGCGTGGGCCTCGAAGAGCGCATGGACGCGAGCGGGGGCCGGG
GCGGAGGCGGCCTGTCTGTCGGCGGCGAGCGCCGCTCCAGTCGTGAGGAG
CAATGCGCGCTCGGCGTCGGAGAGCATCCGGAGCTCGGAGAGCGGTGAC
15 CGGGGTGCTCGACGGCGCTTTCGAGCAGGAGCACGAAGTGGCGGCCATC
CGCTCGATGGTGGCGGGTTCGAAGAGCTGCTGGTCTGACTCGAAGCGCAG
GGCGATGCCGGAGTCGAGCTCTGCGGCGAACAAGGCGAGATCGAACTCGG
CCGCTGCCTGCTCTGTCGGCGAGCGTGGTGAGCTCGAGCTCTCCCTGCGCG
ATCCGCACGTCTCCGCGCCGGTGGTGAGCGCGGCGAGGCGGGGATCCAG
20 CGACGGCAGAGCGCCCTGGAAGGCGAAGGCGACGTGGAAGAGCGCGCCG
CTCGCCGGGCGCGCCCCGGGGCTCTGCGAGCAGGTGCTGGAGGGCGCTG
TCGCCGTGGGCCAGCCCGTCGAGGAACGCGTCGCGCACGCGGGCGACGAG
CGCGTCGAAGGACGCGGCCCCGCGCAGCGCCACGCGCACGGGGAGCATCT
GGACGAAATAGCCGAAAGCCCCGAGTGCTCTCGTCGTCTCCGCCCGCC
25 GAGGGGACGCCCACGACAAGGTCTTCTGCCCCGCTCGCGCGATGGAGCAA
GACGGTGAGCGCCGACAGCAGGACCGAGAAGAGCGTGGTCCCGCGCTCGC
GCGCGAGGCGCGCCAGCGCTCCGGTCAGGGGCTTTGGCAGCGTGATCGCG
TGAGCGCGACCGCGGCGAGGGCTCGCGTCGTGGCGGGCCCGTCCGGGG
AAGGTGATGGCGGTCTGTCGCGCCGTCGAGCGCCTTGCGCCAGTATTCTG
30 CTCCGCCGGCCGCTCCCGCGGCGAGGGACAGCTCACGCCGGCGCGGAAG

5

10

15

20

25

30

35

40

45

50

55

AAGCTCGACGGCGGGCGGCAGCTGCGGGGGCCGGCCCGCGCGCAGCGCCGA
GTACAGCTCCCCAGCTCGCGAACGAGCAGCGCGAACGACCAGTAGTCGA
CCACGAGGTGGTGAACGACCACCGTGAGCAGCGGCGGCTGCCCTCTCCG
CGCCGCCAGACATGCACCCGAGCAGCGGTCCGCGCTCCAGGTGGAACGC
5 GCGGGCGGCGACCTCGTCCGCGCGGGCGACGATCTCGCGCTCGTCCAGCG
CCATCGCCGGCTCTTCGGCCCATTCAGGGCGACATGGCGGTGGACCTGC
15 TGCAGCGGATGGCCGTGCGCGGTGAGGAACGTGCGGAGCGCCTCGTG
CCGCTCGACGAGGCCCTCGAACGCGCGGCGCAGCGCGGCCACGTGACGC
CGGCACCGAGCCGACCGTCTGCCCCAGGTTGTAGAGCGCGCGTCCGGCC
20 GACTTCTGGCACTCCAGCCACATCGCCCGCTGCCCTCGGTGAGCGCAAA
CGGCTCTTCCGGCGTCCGACCCGCGGCGACGAGGGGCCGTCCGGACGGG
TGAGCCCTCGCTCCAGGGCCGTGCTGCGGCGGCGGAGGTGAGGCGGCG
CCGGAGGAGAGATGACTGCGGAGCTGCGCGAGGGTTGGGTGCTCGAAGAG
25 CGTGCGGAGGGTGGTGGGATGCCGAGGGAGGACTCGATGCGTCCGAGGA
CCTGCATGGCGAGCAGGGAGTGGCCGCCGAGGTGGAAGAAGCTGTCGTGT
15 CGTCCGACCGGGTCGAGGTGGAGGACGGATTGCCAGATGTGGGCGAGCTC
GAGCTCGAGCTCGCCCCAGGGGGGCTCGTAGTCGGCGTGC GCGGCGGGGG
GCGCAGGGAGGAGCTTCTTGTCGACCTTGCCCCGAGAGGGACATGGGCAAG
GCGGGGAGCAGGACGAAGTGGGCGGGCACCAGGGCGTCGGGCACCAGGCG
20 GGCCATGCCTTCGCGCAGGTGCGCGTCCGAGGGCGGGTGGGCGTCTGGCA
CGACATGGGCAATCAGGTGCGCGGCGCTGCCTTGGCCGTGGAGGACGACG
35 ATGCCCTCGCGGACGCCGGCAAGCGTCGCAGGACGGATTGACCTCGCC
GAGCTCGACGCGGCGACCGCGGAGCTTGACCTGCTCGTCGCGGCGCCCCG
CGAAGGCGAGCTGTCCGTGCGGGCGCCAGCGCACCAGGTGCGCGGTGCGG
40 TAGAGGCGTGCGCCGGGCTGGCCGAAGGGATCGGGCAGGAAGCGCTCTGC
25 GGTCAGGTCCGTGCGTGTGTAGCCCTGGGCGAGGCACGCTCCGCCGATGT
ACAGCTCGCCGAGGGCGCCGGGCGGGACGGGCTGCATGTGCGGGTCGAGG
45 ACGTAGACGAGGGCGCTGTCGACGGTTCGCGCGAGCGGGGCTCGGCGCC
GAGGTGCGCGATCTCGGCGACCGTGGTGATGACGGTGGCCTCGGTGGGCC
30 CGTACATGTTGAAGAGGCGGAAAGGAGCGGGCGCCGAGCGGATGGAGC

5

203

10

5

15

20

10

25

15

30

35

40

25

45

30

50

55

TTGTCGCCGCCGACGGTCATCGCGCGCAGGGCGGAGCCGGTCCA3TCTTG
CTCGAAGCACGCCTCGGCCAGGGGCGTGGGCATGAATGAGAGTGTGGCCC
GCTGAGCGACAAGCCATGAGACGAGCGCCGTGGGAGAGCGGAGCGCGTCG
TCGTCGGCGAGGAGGAGGGCAGCGCCGAGGCGAGCGGCGTCCAGATCTC
GTAGACGGAGGCGTCGAAGCCGCTGGAGGCAACCTGAGTCCAGCGGTCGC
TGGGCGAGAGATCGAGTCGGAGGTGGAGGAAGGAGACGAGCCTTGAAAGG
CTCGCATGGCGCACACAGACGCCCTTGGGGGTGCCGGTGGAGCCGGAGGT
GAAGAGGACATAGGCCAGGTCTGTCGGGCCTGGAGACGAGAGGAATGTGGG
TGCTGGGCGCGCACGCCCCGCTCCTGGACGAGGTGGACGGGGCAGGGGGCG
GCGGTGAGCTTGTGGCTGGCCTGGCTGCTGGTGAGCGCGAGCGAGGCGCG
GCAGTCGGCGAGCATCTCGGCCAGGCGTGCCGGGGGGTTGGCGGGTCTGA
GCGAGGCATAGGCGGCGCCTGCCTTGAGGACGGCGAGCTGGGCGGCGACC
ATGCGGGGCGAGCGCTCGATGCAGACGCCGACGACGCTGCCGGGGCCGAC
GCCGCGGTTCGCGCAGCCACAGGGCGAGCTCGGTGGACCAGGTGCTGAGCT
GTGCGTAGGTGAAGCGCTGGTGGCCAACTCGAGCGCGGTGGCGTCCGGC
TGTCGAGCGGCGTGGGCCTCGAACAGCGCGTGGACCGAGGCGGGGGCCGG
GGCGGAGGCGGCTGTCTGTCGGCGGCGAGCGCCGCTCCAGTCTGTCGAGGA
GCAATGCGCGCTCGGCGTGGAGAGCATCCGGAGCTCGGAGAGCGGTCTGA
CCGGGGTGCTCGACGGCGCTTTCGAGCAGGACCACGAAGTGGCGGCCAT
CCGCTCGATGGTGGCGGGGTCTGAAGAGCTGCTGGTCTGACTCGAAGCGCA
GGGCGATGCCGGCGTCGAGCTCTGCGGCGAACAAGGCGAGATCGAACTCG
GCCGCTGCCTGCTCGTCGGCGAGCGTGGTGAGCTCGAGCTCTCCCTGCGC
GATCCGCACGTCCCCACGCCGATCGCGAGGGCTGACAGGCGTGCATCCA
GCGATGGCGGGGTGCTCTGGAAGGCGAAGGCGACGTCTGAACAGCGGTCT
CGCTGCGCCTCGCCCTGCGCTCGCGCGAGCAGGTGCCGAGGGCGCTGTC
GCCGTGGGCCAGCGCGTCTGAGGAACGCATCTCGCACGCGGGCGACGAGCG
CGTCTGAAGSACGCGGCCCCGCGCAGCGCCACGCGCACGGGGAGCATCTGG
ACGAAGTAGCCAAAGGCCCTGGCGCTCTCGTCGTCGTGCCGCCCCGCCGA
GGGACGCCACGACAAGGTCTGCTCTGTCGCTGGCGGATGGAGCAAGA
CGGTGAGCGCCGACAGCAGGACCGAGAAGAGCGTGGTCCCGCGCTCGCGC

5

10

15

20

25

30

35

5 GCTCGACGGCGGGCGGCAGCTGCGGGGGCCGGCCCGCGCGCAGCGCCGAGT
10 GGTTCGATGGCGGTTCGTTCGCGCCGTCGAGCGCCTTGCGCCAGTATTCTGCT
CGCCCGGGCCCTCCCGCGGCGAGGGACAGCTCACGCCGGCGGCGAAGAA
15 ACAGCTCCCCCAGCTCGCGAACGATCAGTGCGAACGACCAGTATTGACC
ACTACGTGGTGATCCACCACCGTCAGCACTCTGCGTTTAGTCCTTTCTCC
TGTTTCGGCCTATTAATTGCTACTATGGATCCACACTGCTCCGCCTCTTGT
20 ATCTCCCTTATCTGCACTTGCTGCGCTTNACCTTTACGTTCTCTCCCTG
10 CTCTATACTATTTCTTCCCGCGCTTCTCGTCTTATCTGCATTTGTCATA
TCGTATCTTCATATACCTTTCTTTTCGCTATCCTTACTGCTTCTCGACCTT
ATGTGCGTCTGTCTTCCCTTTCTNTATTATTTCTCTGTCTCACCGCTCTN
25 TGCTCTGTGCTCCTATCACTAAATTATGTCTCTATCACTGCTACTATCT
GAAGCTGATCTTCGAGATCTCGCTNGGTGTCACTCTTTATCTCATAGNCG
15 CCTCTGTCTTCTTGTCTCCTTAAGNCTGATTTTCTCGCTCTATTCTGAC
TACTCTGCTGTCTCTCACATACGTGTTCTTGAATCGTATTTCGCGTTCTCG
30 CTACTGTGATATCCATTGCGGACCTCTACTGCTCNTCTNTATGCTATACT
TCTTAGTCTCTTACTACGTTNGTCTGATATNTTGCTGACGACGTCATGTC
ACGCTCGCAACTCTTCANTTCTATCGTATACGCTGATCATATTTTCTGT
20 GAGGCTGATGTACTATACGTAATTACCTGTATACGTCGTCTATCTACTCT
35 CGTGCTTCACTCTTTCTACTCC

Seq ID No 91 (>Contig12)

40

45

50

40 CCCCCCGCGGTCCGCCGGTACGTGCGGACCGCCGCCCGAGCAGCTCCC
25 CGCGCTCGCGCCGGAGGAGCGGGAGGCCGCGCGCGCCGCTGTGGCCCC
TCGGCGCGGCGCCGCCGAGGTCCGGCGCCGCGGGCTGACGCGGGCGCCG
CTCTCGTACGGGCGAGCCGCATCTACTTCCTCGAGCAGCTCTCGCCCGG
45 CAAGCCGCTCTTCAACGTCCCGGGCGCGGTCCGGCTCCGGGGCCCGGTTCG
ACGTCGCCCCGCTCTCGGCAGCGTTCGGCGAGATCGTGCGGCGCCACGAC
20 GCCCTCCGCACGTGATCGCCAACGTGACGGCGAGCTCCTGCAGATCGC

55

5

205

10

5

15

20

10

25

15

30

35

40

25

45

30

50

55

GCAGCCGCACCGGGCTTCGCGCTCGACGTGGTGACCTCGACGCCCGAGG
AGGCGGCCGAGCTCGACCGCGGGCTGCGCGCCGAGGCGTGGCGGCCCTTC
GCGATCGGCGCGCCCGCTCCTGCGCGCCACGCTGTTCCGCCTCGCGGA
GGACGAGCACGTGCTCCTCGTCACGATGCACCACGTGGTGTGCGACGACT
GGTCGCTCGGCGTGATCCTCCGCGAGCTCCTCGCGCTGTACGCGGGCCGC
TCGCTCCCGCCGCGCGGCTCCAGGTCAGCGACTTCGCGGCGTGGCAGCG
CGAGATGGTCGAGTCGGGGGCGCTCGACGGCCAGCGCGCTACTGGCGAG
AGCGCTCCCGGGGCTGTCCCGGGCGAGCATCTCGGCCGGCGGGGGCG
GAGGCGCCGAGCCACGACCCGTCGCGCGCCATCGAGGAGATCGCGCTCTC
GCCGACAAGGCGGCGGCTCGAGGCGCTCGCGCGCGGGAGGGAGCGA
CCCTGTTTCATGGTGCTCCTCGCGCTCCTCGACCTCGTGATCCATGCGCGG
TCCGGCGCACTGGACATCGCCGTGCGGACGCCCATCGCCAACCGGAACCG
CCCGAGCTCGAGGACGTGGTCGCTCTTTGACGAACACGCTCGTGATCC
GCGTCGATCTCGCGCGCGCCGGGGCGTTCCGCGACGTGCTCGCGCGGGCG
CGCGTCCAGGCGCTCGACGCCTTCGCGAACCAGGACATCCCGTTTCGATGT
CGTACCCAGGATCTGAAGCAGGAGCGCGACCACGCGCAGCACCCGCTCT
TCCGCGTCTGGTGGCGCTCCAGAACGCGCCGAAGCCGCGCTGGAGGTC
CGCGGGCTCCGGTTCGAGCCCCCTGCCCTCCGGCCCGAGCTCGTGCACTT
CGAGGTGCGCCCTCCTGCTCTGGCCGGCGGACGACGGATCGGTCTGGGGC
ACTTCGAGTTCGCGCGGATCGCGTCGACGAGGGCGCGCGCAAGGAGATC
GCGGCCGCATTACGCACCTCGTCGACGCGGTGATCGCCGGCCGACGC
GCCGGTGTGACGCTCGTGAGGGCGCCCGCGCCGAGGCCGCGGAGCGC
AGGCCGCGCTCGGCGAGGCGTTCGCCAGGGCGGCGACGGCGCGCCTCGGC
CAGCTGCGGCGTCTGCTCGGCGGGCGACCGGACGCCCCGCGAGTAGCGGTC
AGCCCTCGGCGGCGGCCAGGCGCACGCGAACGGCGCAGGGTAGCCGTGG
ACGCGCGGCATGGGGTCGATCGCGCTGGGGACGCGGCCCGCAGCAGCTG
CTTGATGGCGAGCGAGATGTGCAGGATGGCCACGTAATTGCCGTGGCAG
TATGGATCCCTGACTCCCAGAACAGTAGTTTCTGCTCGGCAGGCGCCG
GGCCTGAACTGGTCCGGGGCGTCGATGTGCTCGTGGTCTGTCATCGCCGA
GGCGCTGCAGGCCATCACCAGCGCGCCGGCGGGACCTTCTCCTCGTGCC

5

206

10

15

20

25

30

35

40

45

50

55

GCGTGCCACGCCCCGACCGTG TAGTCGCGCACGCAGAGGCTCGTGACGCCG
GTCGACGGGGGACGGAAGCGCAGCGCCTCCAGCACATAGCCGGTGATGGC
GGCGTCGTCTCTCGACGTTACCCAGTTGAGCGCGTCGCGCAGGACGCGCG
GGCGCTTCATCAGCTCGACCAGGGCGTTGACGATCGCGCCGCCGCTGAGA
5 TCCACGCAGCCCATGAGCAGCCCCAGGATCACGTCGCGGATCCCCCTCGTC
GCTCTCGTAGGTCTCGGGGACCGACTGCATGACCAGGTAGCGGTCCAGCA
15 CCGAGGGTTGCTCTGGCGGGGGCGACTTGCCAGCTGCTTCTTCCGCGCG
GCGACGATCGCGTCGATCATCGGCAGCGCCTCCTGACGAGCGGCCCTCGC
CGCCGCCACGGCCGTCGCGGTGCTTGGTGGGTTGAGGAAGATCTCGTTGA
20 10 ACAGCGCGTGGGTCCACGCCACCACCTTCTCGGTGCGGATCTCGCCGACG
CCGAGGTACCGGGCCATCGCGCCGCCGCGCACCTGAGCGCGTAGTCACC
GGTGAGATCGAACGGCTTGTGACGCCGACCTTGGCGAGCAGCCGGTTCTG
25 CCTCGTCCACGACGATCTGACGGTAGCGGGCAGATCGGCGCGCGGGAAC
GCGAGGCGCAGGAGCGACTTCTCGTGCTCGTACTTGGGCGAGTCGTTTAT
15 CGCCAGGATGTTCTGGCCCCACGTTCTCGACCAGCTTGGGCGCGATGTTGT
CGACCGAGAAGACGTCGTTGGCGTTGAGGACCTCGACGACGTCGTTGTAC
30 CCGGTACAGAGCGTGATGGCCGGGATGGAGAAGATGGGCTTCTCGCGCCG
CAGCTGGCTGAGGAACGGGAGCGGCTCCTCCCTCAGCCACTTGAACACCA
TGCCGGCCTCGATCTGCTTCCGCTTGACCGGATCGTTCTCGTGCGCCAGC
20 GCGCTGTGGAGGGCCTGCAGGTAAACGAACGGCGCGCCTTGGCAGCGTC
35 CGCTCGTCCCTCTTCTTCGATGTGAATGCTCATGGGAGAAATTCCTTTCT
CGCATGCCGATCAGATCGCGACGCTCTGGGGGACCATCGACGGGAGCAGG
TACAGGTACGGCTGCTCGCGGGCGCGGTTGCGCTCGGTGATCTCCCGCTC
40 GATGTGCGCCAGGCGCTCGCGGTAGCGAGCAAACGCCTGCTTCGCCGCCG
25 GATCTTGCAACAGGCACTTCATCGTCATCAGGCTCTCGCTGGTCAGCGGC
GGGCCCATGCTGAGCGAGCGGCTGAGCACCATCTGGCGGATGCTCTGCGC
GCGCCCCGGGAGGCGCTCCAGCGGCTTGAATGCCGCTTCTCGCTTCCGT
45 TCAGGACATCGCCGTAGGAGCGGTAGGTGCCAAACTGAGCGTTCGGGATC
CAGGTGTAATAGTCCGTCTGCCCGAAGTTCACGGCGGCGTGATATGCGGT
30 CGCCGTGAAGATGATGTTGGTGACGATCGCGATCAGGTCGTGAGGCTCG

5

207

10

15

20

25

30

35

40

45

50

55

5 TGAGCTTCTCGAGCTGATCGGCTCGCTCCGGCCGGAGGAGGCTATCCATG
10 CCGCCGAGCTGGGGGGACACGAGCTCGTGGATCCACCGCTGCAGGCTGGC
GTGCTCGACAGAGACCCCGGCGTCGGGTAGGCGATCTTCAGCACCTGTC
CGACGTACTCCTGGATCGCGTCCCAGTGCAGCAGCGCTCGTCGCGGTAG
5 TGATAGCCGACCAGGTCGCGGACGTGCGCGCCGACAGGTCGCGGGGGAG
15 CGCGCTCTCGTAGAACCGCCACGGCTTGCCGCGGTACCCCTTTGATGCCCT
TGCCGGTGTAGGCGCGCGTCAAGAGCTCGAACGAGCCCATGGTGCCACC
GAGCTCGTGATGTCGAAGAAGCGCCCTCGCCCGAGGAAGCGCCGGCGAGC
CAGCTCGTTGATGGCCAGGGTGTGAAGAAATGCGGCCTGAGCAGCTGGT
20 10 GGAGCGGATGCGTCGCGGGCAGGTTGCGGTAGGTGCTCACCGCAACGGC
TCCACGATCAGGTGCGCGTACAGCAGGTGGGTCACTTGCCCTGGTAGAT
GGCGTCGGCGCTCGCGACGGCGATCTTCGCGGTGAGCCAGTCGTCCGACG
25 GACCCGAAGGGGTGAAGATCTTGTGCGGATGCGCCCTTTCCCGGGGCGC
GAGTECACCAGCCTGATGGCCACGGGCAAGAGCTCACCGGCCGCGGTCTG
15 GTGCAGCATGCACGTCGGCGCCAGCGGGTACTTGCCAGCTCTTCCTGCA
CGTCGGTGTGACGATGTCTTGAAGATGCGGTAGTCGAGGAAGTAGAGC
30 TGCCCGCCCTCGCGCACCTCCTCCAGCGTGCAGCCGTGCGCGATCGCGAT
CGGCTTGGGCTCGGCGCCGCTCACGAAATCGGCGAGATCGGCCGGGGTTCG
CGCGCGGATGTGCGCCGGGTTGATCCCCACGAGGCGCTGCCGCCGAAC
20 TCGGCGTCCTCGGCCCAGCGCGTCGCCACGAGGGGCTTGCGGATGAAGGT
35 CCACGGCTTGAAGAACTCCTCGAACTGATCGAAGCTCTCCAGTTGTCTGA
TGGACTCGAAGATGGCGCCAGCCGAGGTGCGACGTGGCCCTGAGGACG
AACTTCCCCTCGCGATAGCGCTTGTACCCGTACTCGAACAGGTGAAGCGC
40 CTGCCGATTTGCAGTCCGGCAGTGTCTTCCACTTGCCGAGGTTGAGCG
25 CCAAGATCTTCTTGATCGGAAAGCCTTCCCCCGGCGGTACGGAATCGCTG
CCCGCCGGCAGGTTGGACCCGAAATTGCGCCAGTTCGGTGTGAGCTGGG
CTCCCTCATGCTCGTGCTCGCTTCTCCGTCTCAGACGGACGGTGGATTGG
45 GTGGTTACGTCAAACATCGCTCTCGCGTGCAGCGGTCCGAGCGCGCGC
CGGAATGGTTCCGTCTCAGTCGCAACAGGACTCAGTACATCCAGCGCCGC
30 CCCCCGTCTCGACCTGCCCCCGAGCCGATCGCGCCGCCCTTCATCGTG

5

10

15

20

25

30

35

40

45

50

55

GAATCGACAGGTGCGATTCCACGAAAAGCCGCGCGCCGAGTTGCACGCG
ACCGATGCTCACGCGTGCATTGTTGAGGCTGCTAGAAAACCGTGGAGCGT
TCACGCATGTCAAGCCATTTTGTTCGGCGCCGCGGCGAGCGGCCGATGC
CGCGCGCCCCCGCGCCGGCGTGTTCGCTCCCGACGTACCGCTACCTCGA
5 CGACGTATGGCTTGAAGGGCAACCGCGCAAGTCGTCCGATTCTGTGCTCGT
ATCCTGCTCCTTCCAGCAGGATTTCCCCCGCGCCAGCGGCACAAAGGTGC
CAGGGCGAGCAGAAAGAGCGCTGCCCGCCCCCTCCCCGGCCGCGCTCGCCT
CGATCAGCGCGCGCTCGTCGTGGTGCCTGAGACGTTTCGACGGACGTCAG
TTAGTTAGCCTAGCTAACTTCAACACTGATGCGACTGATCGGGCCGACGC
10 AACCAGCAACCGACGCAACCGACGCAACCGACGCGACCGACGCAATCG
ACGCAACCGACGTGACGGACGCTGGCGACTCGAAGAAAACACGGACGCA
CTCCACGTCATCGACGTCATCGACGTCATCGATGCGCTCGATGCAATCCA
TGCACTTGACGCGATCGGTGCGAGCAGGCGACGAGTCTCTCGTGAAAC
ACCGAACCGAGTGCCGGTAGCGGGCGCGCCGAGTGTATGCTAGGCTCGG
15 CCTCTTGTGAGGCCGCGCGCTCGGCGGTGAGCGTGGGCTCGGGTGCC
GCGGTATCCGGCTGAACCAAGGAGGAGCGAGCCATGCAGGCAGATGACGA
CGCGACGATCTACAAGGTGGTGGTGAACCACGAGGAGCAATACTCCATCT
GGCCGGCGGACCGAGAGAACCCGCTCGGCTGGACGGAGGCCGGAAGACG
GGCAACAAGGCGGAGTGTCTGGCGTACATCCAGGAGGTCTGGACGGACAT
20 GCGCCCCGCTCAGCCTCCGGAAGAAGATGGCCGAGAGCCCCTGAATCGCGG
CCCGCCCGAGCGCCCGTTCGCGAGCGGCCGGCGGGCGGGCTCAGCCGTGTC
ATCGTCGCGCTCGACCGGCCGCGTCCCGCGGATCGCGCGAGCCCGGCGG
GGTCGTGCGCGCCGGCGCTTGTGCCGGGGCCCCGCTCTCGTACGCCTCC
40 GTCATGCCGCCCCCTCGATCTGCACGTGCGCTTGTTCGGCGCCTCCGGCGC
25 CGCAAGACGGTCCTCCTGGCAGCCTTCTACCGGGCGCAGACCCAGCCCT
CGTTCCAGCAGGAGTACGCGTACAAGATCCAGGCGGTCAACAAGGCGCAG
GGCAACCAGCTCCTCGGCCGTTCTATCGCCTCGAAGAGGGCAGATTCCC
45 GGACGGCAGCACGCGCTTCGACGAGTACGAGTTCGACTTCTTCCCGAGAG
ATCTGCCCCGAGCCGGCGGTCCGCACTCCACTGGTACGACTACCCGGGACGC
30 TGGTGGGAGGACGAGCCGGTCGACGCGGACGAGCGGGAGGCGATGCGCCA

5

10

15

20

25

30

35

40

45

50

55

GGGCCTCATCCGGCTCGGGATGAGCCAGGTGGGCATCCTCCTCGCGGACG
GCGCGAAGTACCGGGCCGAGGGCACCGGTACATCCGGTGGCTGTTTCGAG
CACTTCGCCGACGAGTGCAGACCGGCTGCGCCGGGCCAGCGCCGCCACGGG
CGACGAGGTGAGCTTCCCGCGGGAGTGGATCCTCGCCCTCAGCAAGGCCG
5 ATCTCTGCCCGCCGGACTACAGCGCGCGGGACTTCGAGCGCGAGGTCTGC
CGGGACCCGACGATCAGCTGGCGAAGCTCTGCTCGGTGCTCCGCGCCGA
15 GCACGCGTTTCGGCCACCGCTTCATGCTGCTCTCGTCGGTCGCGCCCCCG
CCGGCGCGCAGGTCGATCCGAGGACCTCGCTCGGCGTGCGCACCTCGCC
CCCGCGATCCTGGTGAGCACGGTCGAGGGCGCGGTGCGCGAGGCGCAGGC
20 GCGGAGAAAGGAGAAGTCGCGCGGAGAGACGTTCTTCAGGGGCTGCCCG
10 ATCTCGTGAGTTTCGTCGACTCCCTCGACGACTTCCTGCCGAAGCGATAC
CAGATCGTGAGCAAGATCCTGCGGTTTCATCTCGATCAAGGACTTCGCGAC
CACC CGGCTCGACCGGCTCAAGAAGATGCGCGAGGACGCGATCCGGAAGG
25 GCGACACCTTCACGGCGGTCTTGACCGCGATGGTCGCGGCCCTGCGCGAC
GACGAGGGCGCCCGCGCTACCACCAGAACCAGTGAGGTCGTCATGCCCG
15 CGCCAGCGCCCCCTCGTCGAGACATCGCGCCTCCTCTGGAGGACGCGCGGC
CGCACTGGGATTACGAGTTTCATCTGTGTCCCGAGATCCCGGCGCTGCC
CGCTGGCTCTCGACGCTCGAGGCGATGCTCGCCGACGCCGACGCCGGCG
30 CCGGGGAGCTCCGCTATGGCCTGCTCGAGATCGACGATCGCGGGCAGAGG
CGCCCGCGCGCCTATCCCTACGTGSCCGTGAGGTTCTCTCGATCCGGCGCG
20 GAGGGACTGGACCGGACGGCAGGTCCAGCACTTCGCGGCCTGGTTCCCGC
CGGTCCCGCCCGAGGCGGTGCGGAGTTGCCAGAAGCGGTCCCCGCCGAC
TGGCACCTTCGCGTGCTCGACGGCTCGCGGGACGTACGGCTCCGGCGA
40 GGTGTTCCGGGCTCCCCGAGGCGACGATCCGCGCCTGGAAGCGGAGCCACG
25 ACGAGAGCCGGGCGCGCGCGATGGCGATCGTCAAGGCGACGCCGCCG
GTTTCGCTGGGCGGCGGCGAGGCGCGCCGTCGCGGTGGACGCGGGTGCC
GACATTAAAAAAAAGCCGCCGGA3CCGCCGGCCGCGGGGCTCCTCT
45 CGGTGGGCGCGGTCCCTAGCGGCCAGGGCCGGCGATTGCGTGTCTCGCG
ATCGGCGCATGATGCTCGCCGCTTCTGTGACTGATGCTCGCTTGCGG
30 TGTGCGCCTCCTCGGCGCCTGACG3CTGCGCCGCGCAGGCCATCCGACGG

5

10

15

20

25

30

35

40

45

50

55

GGGGTCGGCCCGGCCAGCGCCCGCCGGGCGACACCAGGGCATCGGCCCTC
CGCTCGGGGCATCGATTGAGCTCTCCGAGCGGCGGTCCGTCGTCAATCGC
CGCAGAGCTCCCAACGGGCGGAGCAGCTCTGGCCGGTGACCGCATAGGGG
TTCGTCGGGCAGGTCCACCACTCGCCCTGGAAAGGACGCGGGTTGCAGTG
5 CGGGAGGCACTCCACCCACCCGACGAGCACGAGTTCCTTACCGAGACGG
TCGGCTGAGCCGCGCAGAACCAGCGTTTTCTTGCGAACCAGCCGGGATTG
CACACATCGGGCGCCCCGCCGACGGGCGGATACACCGTGCGACCGCCTG
GATGTCCACGGCGTCGAGGGTCTCGCTCCCGAAGCGACGCCGTCCTGAT
CCGAACACCCACTCCAGTAGGTTCATGATGGAGTCGTAGTCGTAGAAACCA
10 GGGTTACGACGATGTACCGCCGGCTCGAGGGCCACCCGCTGGCGACGTC
GCTCGCGGGGAGCGGCTCTCGTTGGCTGCAGGCGCTCGGGACCAACGGGT
GATGCCACTCATGCATGAAGCCGATCGCATGACCCATCTCGTGATCGCG
TACTGCTCCACGCAGTCGAAGCTGTATTGACCCGGGCTGTCTGCCAGTT
GTACTTTGATGCAACGGTTGAAGTCGGCGCCCCAGGGCTTGAAGTGGACCG
15 AGCCGCCCTTGTTGTAAACGCCGATCGAGTCCGATTGGTTGGGCGCGTCG
GGGTGGATCCTGACGCCGACGTAGGTTCATGCGAGTGGCCGGCAGGAGCGA
ATCGCAGCTCTCCAGCCGGTGAAGCGAACCGAGCTCCAGCGTTCCAGC
TGCCCTGGAGCGCGGTGCGCACGCGCGTGATGACGTCCGCGAGCGAGGGG
TTGGGCGCATGGATCAGCCCGCCCGCGGCGCCGTCGACCCTCTGCTCCGC
20 CGAGCTCGTGGGCTCGATGCAGACCGGGATCCGGACATGGCCGTCAGCGT
CCTCAGGCCAGCGACTCGCGCTGTGGAAGACGCTCGCCTCGGCGGACCGC
GGCGCGGCAGAGCGGTGACGCGGGCGCCAGCGCTGCGAGGAGCAGCGG
ACCGAGCGAAGAGCGAAACCGCACATGTCGTTTCAAGGCCCCGCGTCGTGC
GGTGACCCGAGACAATCTCGAGCGGGCTCATGGACGCAAACGCGTTGCGA
25 TGGCCTTGACGATGTTCTTCTCCAATCGACGAGGGTTGTTCTGCTGAAC
GCGGCTCCAGCGTGGAGCTCGACGCGGTTACCGGCTTACGCCGGGGCC
GTGGACGAGACCCGAGCACGGGGGAGGTCCGAGCCGCACCGGCTCGCGGC
GCCTCCACCCTGCACCTACGACGAGCCTGCCGCTCGGTTTCGCGGAAAAT
GCCACCCCGCTGCCCAGCGGGCGAAGCGCGGACGAGGCGCTCGTCCCCAC
30 GGTAGCGCCGGTGCCGCTGCATCCACCGCGCTCCTCCATGGGTCGCTGCC

5

15

20

25

30

35

40

45

50

55

[illegible]

5

10

15

20

25

30

35

40

45

50

55

CTCAGCGCCGGGGCCTACCTTATCGCATCTTGGGCGCTTGGCGTCCAGGA
TTCGGCCTTAGACAGCACAAAGCAGAAGACCTTTGACACTGGATTTTTTCA
TCATCGGCGGCGCTCGTTCTTCGCTGCGCCTCAAGCGCCGACCGTTTCGTT
TCGAAGCGAAGCGGTTTACGATCCGGGATCGCGGAAATTTGAAACGGAC
GCGTCGCGCGGGCAACGCAGGGGACTCATCACGAGGCAACCGCGCTGCGT
CGCGAAATTGGCCAGCCTCTCGGAGTCCCTAGTTCCGTGCGTCAGACGCC
TCACCCACCATGTCGAGCTCGGCGCGGCGCTCTACGTGCTTGAAAGACCT
CCGCGAGCGCCGCGCTCTGCGCTCCGCGAGCGCACAGGCTCCCGTGGA
TTCAGGGCAAGGCGGTCTGTGATCACGTCCTTCGTGGTCGCCGTGCTCTGG
CCGGGGCGGTTCGACGCAGGACGGGTAGTACTGGCCGGCCCCGTTGAGCTT
CTGGCAGACCCGGTCGCACGACCCGGCGATGCGGATCAGCCCGCACTCCA
CGATCTGACCCCCGGAGGTCACGTGGCCCGCGGCGCAGTCGCGCTGGTAG
GCGCGCGAGTTGTTCGACCGTCGCGCTGTTGTAGCAAGCGTTGATATAGGG
TTGCGCGGCGAACAGGTTGCCCCAGAAGGCCCCCTCGACGTCCGGATAAT
CGATGAGCTCCTGGCTGGAGGAGAGCGTCTTCAGCGGATCCCGCAGCGAG
CGGGCGGAGAGGAGCACCGGTACTTGATAGTAGTTACGCGCGCCGCCAC
GCAGCTGGACACGATGCGCTGCCCTGCGTCGTTCGAGCGGCCCCGCTCGCCC
ACGCGGGCGCGACGCCGAGCAGCCCGGGGTAGCGCTCGTCGTGCTCTTG
CCGTTTCGAGTCCGTCCACGAAAAATCGAAGGAGGCCGTGCTGCTCAGGGC
GCAGCTCGCCGCGTAACGCAAGAAATCGCGCGCCAGCGCGCCGCTCGGCC
CGGGATCCTGGATCGCGGCGAGGTTCCGCGCGCTGAGGCCGCTCAGGTTT
AGGGCGTTTCAAGTTTCAAGGCGTTGAGGTTTCAAGGCGTTGAGGTTTCAAGG
GTTTCGTGCTGAGCGGTTGCGGCCACGAGGGCCCCCTGGGATTCCCCCA
CAGGCTCGCCCCACGCATCGGCGTCCACCACCTCGGCGGCGCAGCCCGAC
AGCACCCTGCCCAACCAAGCACGATGAATGTCCGCTCGAGAGACATGGA
TTCCCCCGTGTTCCTGGCGCATGACCCGACGGCGCCCTGCGCGCGGCGCG
CGCGGGCTCCCATCGATTGCTGATGGGTTCAATATTCTACTTTTCCC
GCGCTCTCGCGCCGGTGAAAGTCGCTTCAGCGGCGGCGAGGTCGATGTCA
GGAGCGTCCGACTCCGTCGCTCTCGTCAGCTCCGCGTACCAGCGACGGAG
TCGCCCCCGCATGACGGTCGGAATGGTAGAGGCGGCGGAGGGCGCGCT

5

213

10

5

15

20

10

25

15

30

35

40

25

45

30

50

55

CGAGCTGCGCCCGGGCGTCGGCGCGCCGGCCGCGGCGCAGGGCGGCGAGC
GCCCCGCGCTCGAGCACCTCGATCCGCTCCTGCCCCGACCGAGCAGCGCGC
GGAGCGCTCCTCGAGCGCGGCCACGCGGCGCGGTCTGTCGTCGCGGGTCG
CGAGCTCGATCATCGCGCAGAGCACGTCTCCGAGGGCTTCAGCGCCTCG
CAGCCGGCGTCTGTCGCGCCGCGCGGAGCCGAGCGCGATCCGGCGAGC
GCCCCGCTCGTCGCCC'GGTAGAGCGCAGGCGCGCGATCAGGAGCGTTA
CGACGACCGGCGCGTGGCGATCGCCGAGCGCGGCGCTATCGCCTGGACG
GCGCGCGCATGGGGCTCGCGGCCCGAGATCGTCCATCAGGTACAGGTA
CTCGGCGAGGTTGTAGCGGCCGACGAGCTCGAACGCGGGCTGGCCGAGCT
CGCGCCCCGAGCGCGATGGTGCGCTCGAAATCGGCGATCATCCCGGCGCGA
TCGCCCTGGAGCGCCCCGCGGAGCCGCGGTTGTTGAGCGCGGCGCCGAG
GTGCATGAGATCGCTGCGCTCCTCGCAGCTGAGGATCACCGCGTCGAGGT
CTCGCGCCGCTCCTCGACGCGGCCGAGGCTGGCCAGGATGAAGCCGAGC
AGCAGCAGGGCGATGATGTGCGTCTCGTGGCCCTCGTCCCCGAGCCGCGC
CGCCTGCGCCCGGCGCGCGTCAACACCGCGCGCCTCGTCTCGCGGT
CGGCGCGGTGGAGCGAGCGGCCACGCGGAGGAGCAGGCGGGCGCCGAGC
AGGGGCGAGGCCACCCGCGCGGCGAGGCGCTCGGCGGCCGCGACCCGCTC
GCGCGCGGCCCGGTACTCGCCCGTCCAGTCGAGGATCATGGCCTCGTCGA
GGAGGAGCTCGATCTCGGCCCCCGCTCCGACGCGCCGCGCCGCGCTCG
CGCGCCGCGGCGAGGTGCGCGAGGCGCTCGGTGTGGCGCCGAGCCGGAA
GCGAGCGAGGCCCGCGCTCGGCGCTCCTCGGGGAGCAGCGCGCCGAGCA
GCGCCTCGACGCGCCCGTAGCAGCCCTCGGCGTCGAGGTAGGCCCGGCGC
GCGGCCGCGAGCTCGGCGCCGCGGCGAGGAGCGACGCGCGCGGGCGGT
CAGGCCGCGCGCTCGCAGTGCGCCGCGAGCACGCGGATCGGCCTCGC
CCGCGGCCTCGAGCCAGTCGGCGCGGAGGCGGTGGCCGAGCGCGCGATCG
GCAGCTCGCCGCGTAACGCAAGAAATCGCGCGCCAGCGCGCCGCTCGGCC
CGGATCCTGGATCGCGGCGAGGTTCGCGCGCTGAGGCCGCTCAGGTTC
AGGGCGTTCAGGTTCAGGGCGTTGAGGTTCAAGGCGTTGAGGTTCAAGGC
GTTCTGCTGAGCGGTTGCCGCCACGAGGGCCCCCTGGGATTCCTCCCA
CAGGCTCGCCCCACGCATCGGCGTCCACCACCTCGCGCGCGAGCCCGAC

5

10

15

20

25

30

35

40

45

50

55

AGCACCCCTGCCCAACCAAGCACGATGAATGTCCGCTCGAGAGACATGGA
TTCCCCGTGTTCCCTGGCGCATGACCCGACGGCGCCCTGCGCGGGCGCG
CGCGGGCTCCCATCGATTTCGCTGGATGGGTTCAATATTCTACTTTTTCCC
GCGCTCTCGCGCCGGTCAAAGTCGCTTCAGCGGCGGCGAGGTCGATGTCA
5 GGAGCGTCCGACTCCCTCGCTCTCGTCAGCTCCGCGTACCAGCGACGGAG
TCGCCCCCCCATGACGGTCGGAATGGTAGAGGCGGCCGCGAGGGCGCGCT
15 CGAGCTGCGCCCCGGGCGTCGGCGCGCCGGCCGCGCGCAGGGCGCGAGC
GCCCCGCGCTCGAGCACCTCGATCCGCTCCTGCCCCGACCGAGCAGCGCGC
GGAGCGCTCCTCGAGCGCGGCCACGCGGCGCGGTCTCTCGCGGGTCG
20 CGAGCTCGATCATCGCGCAGAGCACGTCCTCCGAGGGCTTCAGCGCCTCG
CAGCCGGCGTCTCTCGCGCCGCGCGAGCCGCGAGCGCGATCCGGCGAGC
GCCCCGCTCGTCGCCCTGGTAGAGGCGCAGGCGCGCGATCAGGAGCGTTA
25 CGACGACCGGCGCGTGGCGATCGCCGACGCGGCGCTATCGCCTGGACG
GCGCGCGCATGGGGCTCGCGGCCGCGAGATCGTCCATCAGGTACAGGTA
15 CTGCGCGAGGTTGTAGCGGCCGACGAGCTCGAACGCGGGCTGGCCGAGCT
CGCGCCCGAGCGCGATGGTGGCTCGAAATCGGCGATCATCCGGCGCGGA
30 TCGCCCTGGAGCGCCCGCGCGAGCCCGGGTTGTTGAGCGGCGCCGAG
GTGCATGAGATCGCTGCGCTCCTCGCAGCTGAGGATCACC GCGTCGAGGT
CTCGCGCCGCTCCTCGACGCGGCCGAGGCTGGCCAGGATGAAGCCGAGC
20 AGCAGCAGGGCGATGATGTGCGTCTCGTGGCCCTCGTCCCCGAGCCGCGC
35 CGCTGCGCCGCGGCGCGCGTCAACCGCGGCGGCTCGTCTCGCGGT
CGGCGCGGTGGAGCGAGCGGCCACGCCGAGGAGCAGGCGGGCGCCGAGC
AGGGGCGAGGCCACCCGGCCGGCGAGGCGCTCGGCGGCCGCGACCCGCTC
40 GCGCGCGGCCCGGTACTCGCCCGTCCAGTCGAGGATCATGGCCTCGTCGA
25 GGAGGAGCTCGATCTCGGCCCCCGCTCCGACGCCGCGCGCCGCTCG
CGCGCCGCGCGAGGTGGGCGAGGGCTCGGTGTGGCGCCGAGCCGGA
GCGAGCGAGCCCCGCGCTCGGCGCTCCTCGGGGAGCAGCGCGCCGAGCA
45 GCGCCTCGACGCGCCCGTAGCAGCCCTCGGCGTCGAGGTAGGCCGGCGC
GCGGCCGCGAGCTCGGCGCCGCGGGCGAGGAGCGACGCCGCGCGGGCGGT
30 CAGGCCGCGCGCTCGCAGTGCGCCGCGAGCACCAGCGGATCGGCCTCC

5

10

15

20

10

25

30

35

40

25

45

50

55

CC3CGGCCTCGAGCCAGTCGGCGGGCAGGCGGTGGCCGAGCGCGGATCG
TCCTTGGTGAGCTGCGCGTAAGCGCCCTCGCGCAGGAGCGCCTGGCGGAA
GGAGTACTCCTCCTCGCCGGGAAGCGGCCCTCGCGGTGGCGGACGCAGA
GCTCCCCGGCGACGAGCGCGGAGAGGTGCTCCGCGAGCGGAGCGGCCTCG
TCGCCCCGAGCAGGTGCGCGACGGCGCCTCGCCAGAACACCTCGCCGAG
CACGCTGGCGGCCCGCAGGATCCGGCGCGCGGGGGCGCGAGCGCCTCCA
GCCGGACCTGCACCATCGCCACCACCGTCTCGGGCAGCGGTGCGCCGCG
CCCTCCGCCGTGCGCGGATCAGCTCCTCGAGGAAGAACGGCTGGCCCTC
GGACTGGGTGACCAGACGATCGATGAGGGCCCCGTGCGCCGCGTCGCCCA
GCGCCTCCCGCGCGAGCTGCGCGCACGCCCTCGGCGGGAGCTGCCTGAGC
CAGAGCTCCTGCCGCCCCGCGCTCGGCCAGAGATCGGGGTACGCTTGC,

or their complementary strands,

(b) DNA-sequences which hybridise under stringent conditions to regions of DNA-sequences according to (a) encoding proteins or to fragments of said DNA-sequences,

(c) DNA-sequences which hybridise to the DNA-sequences according to (a) and (b) because of a degeneration of the genetic code,

(d) allele variations and mutants resulting by substitution, insertion or deletion of nucleotides or inversion of nucleotide segments of DNA-sequences according to (a) to (c), wherein the variations and mutants offer isofunctional expression products.

13. Peptide encoded by a DNA sequence according to claim 12 selected from the group consisting of

5

10

Seq ID No 92

>Contig11_002 591 amino acids MW=63639 D pI=5.80 numambig=0
MLDVWSTSDQVACRLHCAGAGPSASLELRYDASAGARRDAERLAERLAALLEDLSRHPER
5 PVAQGEYVGPGERAEIEAWSRGPAMELPSACALHRWFEERAEQHPDVVAVRSEGKSLTYG
15 ELERRANRLASCLRRRGVGLDTIVGVCVPRSEDMVVATLAVLKVGAYLPDHEYPGERL
AFMMRDARARLLVTHDAIADELPTGGWTTLLLDAAEAEIAACSDARPAVSPPPDSGAYVI
YTSGSTGTPKGSLSHRAIVNQMWIORYWALTADDRVLLKAAFQFDVSVWEIFWPLSFG
ARIVVARAGGHRDPEYLRRLVRDEGATTAYFVSSMLAAFLGGPEQFPASLRKVLVGGEA
20 VPLDLVRRFYAKHGDGLINMYGPSEAAIAVTGCVLPSPRVTWVPLGAPVANAEEVFLDG
10 AMRRPAIGALGDLYIAGAPLARGYVGQPLTAERFLPDPCARAAGGRMYRTGDVARFLPD
GMLEFQGRSDHQIKLRGHRIELGDVEAQIRRVPGVGQAQAVVLREDAPGDARLVAYVVLGD
25 DAAGDAPDVRAGLKASLSAYMIPSSVRLYALPMCSERLAFTGSSYAGCLL*

25

15 Seq ID No 93

>Contig11_007 361 amino acids MW=38862 D pI=10.42 numambig=0
MSDHMTGFSLSPQQRAIRALDREAGAPGCRTLAVVAVTGPCDEGRLSAAALALAERHEI
30 LRTRLVEAGRARPRRWSASRASRGRQQDDWVCSEAEQGERMSRLVARLSEDRGADDGLR
VGLVRVGPEERRLVLAAPAWCVDEESIAPLVRELCASTAGAGAPPEQQYADVAEWLNGML
20 ESEDAGDGRRFWAERRSHFGPPLHLAFSRGGAGAGAGSGRARVDLGQGGMAQVERWSSSW
35 QVPQRIVLLALWASLLWRMSGGNEPEVTVAVRFDGRSLDALAGAVGPFARFLPVRIEISA
SDTLADVARRLALAEAEAAAHQDAAPGVSHRMSWGLLRGGRAGAVARRRAGPRARRLEH
V*

40

25 Seq ID No 94

>Contig11_012 882 amino acids MW=95015 D pI=12.69 numambig=0
MARALYAQEAARSAFNACAAAARAGLDLGEVVFREGECDGRTLLGTALAQPALFAVEL
45 ALARLWMSWGIEPAAMIGHSLGELVAACVAGVFTLEDAMSLVIDRGRFMQAAPAGSMLAV
GLPAADVEGLLEAGLEIAAVNSPKLTVVAGPASAIRDLAARLEAREVFARPLQTSFAFHC
30 ALIDGAVAPFLESVRRARLSPPEIPVVSNTGALLTDAEATDPAYWARHLRQPVRFSDGV

50

55

5

10

15

20

EALFASCHALFLEVGPGRCLTTLVRQTLACRGGAIASLGSTHAASEPASLAEALGQLWE
AGHAVDWTARPRGRPPARSSACRRTRSSGRGTGSRRAAAPPPPSRRRPTRPSRPRRTPSR
RRRSPARTPGPRSPRRRPSPSARS PAIWERPPRRPRDRRPRRLLRPARRLARRDPGDVPD
PRAARRRAPAAGALPGADGRGARGPGRRRAPRRGAAARVPPDRADPPRRPAPAVVRAAPA
5 VVRRSARAGQPGVQHPVRGARDGPARRRRAPPQPVRDRAAPRGAADDVQRQGRRAVPGRR
ARGPGAVPDVRPRAPRGRGPRRGGLGARPRGEPRAVRPEPRAAPPRAGHPEAPRRRARHRA
15 RRAPRRVRRLVGRRLRRRARRALRRLRAGPAFPAAGAPGAVRRLRRSPAGVAVRRGPGR
APVLDDEALGRAAPGPRAGRPRAGGAKDVARGAAEPRRRGRAHPPDQGVLRAGGDLPLHG
PARRVQARPAPAHRRARGPRRGHRRREPEPRRDADDRLLRQPARAPDRLRGRPDVRRAGP
20 PRARRRRARGVRAPGPAVRPARRGPAEEGRRRAPLRREVRHAQARPPDEARGARARGA
10 GGRGDHDGVRLRPHGRRGRRLVPLRRRAQRAVPGDGRQFP*

25

Seq ID No 95

>Contig11_021 1213 amino acids MW=131017 D pI=12.40 numambig=0

15

30

35

40

45

20 MRGRRRRRAAPHLRGARPARRGDRRGGLALRSDREARRRGHGEGLGAGGRRARRAARRRR
LPAPGPSPPRGAPQAPPGARRGPPRPHAVGGRRDDRLARGHRAPRRRRRGRAVARAARRAP
AARRKHGRPGLRDLHVRVHGAAQGRDDRPPRRGEHGPRHQPLRRRPPGGPGARALVAELR
PVGLRRVRDARRRRRRRDPRPYRLGSGALARARGARAGDRVELGPGADGDAHGRVPRRG
RPGAVVAPPRHDERRLDPAEAPRSHPRGLPRAPRREPRRRDRGVLDVDRPPDRGRRPVGA
QFPLRPPAREPAHLRARRGAGAVSDRGPRRDPHRRDRRRARILARRGADPGAVPEAPHDR
RAAVQDRRPGPLRRRDHRAARAHRPSGEDPRLPHRARRDRGRPRAAPLGRAGGRGGEDR
SVGREAPGRVRRRRRRRRRAPRLRPEEAAGVHDPRGGRPPGAAAERERQGGPRRPAGS
RRGRAARRGRRPADRDRAAHLRGAGGLAGRGRRRDROPLRARLHLAAPRARAAPARRAH
RGARARRGRRGAGRVADRPVVPDHRATGAAARRGDGQGRAGRRRGAARRGAPRRATPPG
25 ARIAMSEPIETEDGGS DIAIVGMAGRFPGAPSDALWENVRRGVES IARFPESEREPPV
GASAAPGAPVVCAGGLDDIDRFDASYFGYSPREAQLMDPQQRLFLECAVAALDAGCDP
ARFPGAIGVFGGCGSNTYLLQLLSHPDLAATVDPHALMLASEKDYLATRVSYKLDLHGPS
VVVQTACSTSLVAVHMACESLLGGQCDLALAGCVSIGIPQKRGYPYVPGSICSPDGRCRP
FDARAEGTVGGSGVGIVALKRLAJALDRNTVHAVIRGS AVNNDGGRKVGFMAPSVDGQA
30 AAISEAQSVAGVDPG SIGYVEAHGTATAIGDPIEVEALTQAFRRKTPRKAYCALGSIKAN

50

55

5

218

10

IGHLDAAAGVAGLIKAAHVVRSGEIPPCVHFEAPNPKLDLAASFFVPREAAPWPRELRP
RRAGVSSFGIGGTNAHVVEEPPPLPPRAPAPERDHLVTLSTARTPEALSTACAQLAAHLE
ATDVPLDDVAFTLTQTGRAEHYPYRAVVARTRAEAIQGLAREGASALARPDEPRPSSRSRA
RARRPSGWPASTRRRRRSGAPSTRARRRRGRAASISARSSSARARATGARCSAPRWRSP
5 RSSPSSSRSPGSG*

15

Seq ID No 96

>Contig11_026 3079 amino acids MW=332984 D pI=5.97 numambig=0

20

MLTVVDHVVVEYWSFALIVRELGELYSALRAGRPPQLPPSSFFAAGVSCPSPREAAGG
10 AEYWRKALDGATTAIDLPRDRARHDAGARRGRAHAITLPKPLTGALARLARERGTTLFSV
LLSALTVLLHRASGQSDLVVGVPASAGRHDDESARAFGYFVQMLPVRVALRGAASFDALVA
RVRDAFLDALAHGDSALRHLLARAQGEAQRDALFDVAFQSTPPSLDARLSALAIGVGD
25 VRIAQGELELTTLADEQAAAEFDLALFAAELDAGIALRFEYDQQLFDPATIERMARHFVV
LLESAVEHPGRPLSELRLMSDAERALLDDWSGAAAARQAASAPAPACVHALFEAHAARQ
15 PDATALEFGHQRTYAQLSTWSTELALWLRDRGVGPGSVVGVCIERSPRMVAAQLAVLKA
GAAYASLDPANPPARLAEMLADCRASALTSSQASHKLTAAPCPVHLVQDGACAPSTHIP
30 LVS RPDDLAYVLFTSGSTGTPKGVCVRHASLSRLVSFLHLRLDLSPSDRWTQVASSGFDA
SVYEIWTPLACGAALLLADDDALRSPTALVSWLVAQRATLSFMPTPLAEACFEQDWTGSA
LRAMTVGGDKLHPLRRPLPFRLFNMYGPTEATVITTVAEIADLGAEPPLGRPVDLSALVYV
20 LDPHMQPVPPGALGELYIGGACLAQGYTRTDLTAERFLPDPFGQPGARLYRTGDLVRWRP
35 DGQLAFAGRDEQVKLRGRVVELGEVESVLRRLPGVREGIVVLHCQGSAAHLIAHVVPDA
HPPSERDLREGMARLVDPDALVPAHFVLLPALPMSLSGKVDKLLPAPPAHADYEPPSGE
LELELAHIWQSVLHLDRVGRHDSFFDLGGHSLAMQVLGRIESSLGIRTTLRTLFEHPTL
40 AQLASHLSSGAASTSAAAATALERGLTRPDGPSSSRVATPEEPFALTEGQRAMWLECQKS
25 ADGALYNLGRTVRLGAGVDVAALRRAFEGGLVERHEALRTTFLTRDGHPLQQVHRHVALEW
AEEPAMALDEREIVARADEVRRRAFDLERGPLLRVHVWRRGEGQPPLLTVVVHHLVVVDYW
SFALLVRELGELYSALRAGRPPQLPPSSFFAAGVSCPSPREAAGGAEYWRKALDGATTA
45 IDLPRDRARHDASPRRGRAHAITLPKPLTGALARLARERGTTLFSVLLSALTVLLHRASG
QNDLVVGVPASAGRNDDESTRAFGYFVQMLPVRVALRGAASFDALVARVRDAFLDGLAHGD
30 SALQHLLAEPRGAARRGGALFDVAFQFQALPSLDPRLAALTGAEDVRIAQGELELTTL

50

55

5

10

15

20

25

30

35

40

45

50

55

ADEQAAAEFDLALFAAELD SGIALRFEYDQQLFDPATIERMARHFVLLLES AVEHPGRPL
SELRMLSDAERALLLDDWSGAAAAARQAASAPAPACVHALFEAHAARQPDATALEFGHQRF
TYAELSTWSTELALWLRDRGVGPGSVVGVCIERSPRMVAAQLAVLKAGAAAYASLDPANPP
ARLAEMLADCRAALVLTSSQASHKLTAAPCPVHLVQDGACAPSTHIPLVSRPDDLAYVLF
TSGSSGTPKGVCVRHASLSRLVSFFQHLLALSPDRWTQLASSGFDASVYEIWTPLACGA
ALLLADDDALRSPTALVSWLVAQRATLSFMPTPLAEACFEQDWTGIALRAMTVGGDKLHP
LRRPLPFRLFNMYGPTEATVITTTVAEVDLGDPEPLGRPIDSALVYVLDPHMQPVPPGVL
GELYIGGACLAQGYTRDTLTAERFLPDPPGQPGARLYRTGDLVRWRPDGQLAFAGRRDEQ
VKLRGRRVELGEVESALRRLPVAVREGVVVLHGQGSAAARLIAYVVP GADPPSERDLREGMA
RLVPDALVPAHFVLLPALPMSLSGKVKLLPAPPAAHADYEPSPGELERELAHIWQSVL
HLDRVGRHDSFFDLGGHSLLAMQVLGRIESSLGIRTTLRTLFHPTLHQLADRLSSGAAS
TTAAATVPASEIAPSLGRAPADEPYPLSYEQERLWVLEQLLPGGTAYNVVQAVRLRNLV
DVDALSSALAAALVRRHWSLRTVTFVASPTPAQKICEPEAAPAEVVDLRGTPPDEAEAAARA
WASREQATGFDLARGPVFRARLFRLDHDCVVLVLSHHIVTDAWSFQPLVRDLAELYRRA
RGGGPADMPPELPLQYVDFAVWQRRHLAGKRLADKLAHWATLRLGPLVLELQTDPRPPVQ
TFRGAERVLPLDARLVAQLDELARSRGATR FMVLLAALGVLLRRSSGQDDLAIGTAVANR
PRPELEPLVGFFVNTIVMRDLGGDPTFEELLSRARKVALEAFEHQDAPFEKVVEAVNPR
RDLRSRPLFQVMLVVQNAPTEALELGEVRIEPLDLPVEATRFDLRF SVEPRGGRDVISLQ
YNVDLFDAAATIDRMLATMQSVLSRATQDPAQVRALSVAPEDRERALVAWNDTAVATPDH
LRLEEPFFFERAVEQPDACAVVDAERRLTYGELARRAEIAAAAASRSGATANALVAVVMEK
GWEQVAAVLVGLRAGAAAYLPLDPRLPEERLRHLLHAEVRLVLTQSAVDGTIAWPAGIER
LAVDADERWREQPVARRPPGGSTDDLAYVIYTSGSTGLPKGVMIDHRGAVNTVLDINRRF
DVGPEDRV LALSSLSFDLSVYDVFGT LAAGGAVVIPDRTRASDPGHWR ELVERERVTVWN
SVPALMEMLMDASPGAGDPALSSSLRLVMMSGDWIPLKLPDRIR AACRAPRVVSLGGATEA
SIWSIAHPIADVDPAWRSIPYGRPLANQHTYVLDEGLEPCPIGVPG EIHIGGIGVALGYW
RDEARTRERFLKHPTTGERLYRTGDLGRYFADGTIELLGRTDHQVKIRGFRIELGEIEAA
LAQHPSVEQAVVAAKTDPSGEKRLVAYVVGADGDGAALRDFVRKKLPEYMI PAEVVVLPA
LPLSANGKVDRAALPDPAAVAPRAAAVAPRTATERLIASVLAEVLQVEAVGVTDNLFELG
FTSLLLVR AQRLLAERIAARAPDEGAAAQAVSLTDLFQYPTIEQLAQRLDAATVKAEPAD
VGAQRAEARRDARRRRGRG*

5

10

Seq ID No 97

>Contig11_011 544 amino acids MW=60164 D pI=9.10 numambig=0
MMSRIRAQLGVLEPLRALFQGPTVAALAAQVDAARRGEARRREFPPIARIPRDGGLPLSF
5 AQHRLWFVDQLEPGSPAYNIPFVVRATGRLDVDALRRSLFEIARRHEALRTTFSARDGVP
15 FPVVAPEARVPFRMSDLEHLAGEALDAAVSALVLEESLAPFDLSRGPLLRVRVIRKRHDE
HVIALVVHHVVDVWSVGVFVGELAAALYGGFAQQQPSRLPELPAQYVDFAAAQRAWLSGE
VLEGELRYWTTKLSCALRRARVPVDHEPAGRRTWRGARRSLDAGAELTRQIKAFCEAREAI
SPFMALLAAYKLVLHQRTGLEDLVVGTDVANRRNVETEPMIGFFVNQLVLRDCCGGDPTF
20 GALVRRVRDVALEAFEHQDLPPDRLVEALRPKGAVGHVPLFDAKFVMNRNVHVPMMKLEGL
10 ELEALEGEATTTAFDFVLTVAEAGGSFRFGVEHSSSELYRAATVDNFLSDYRQILATATAR
PDTPVSELRGELERAAAARELERKAARGAALDKLTSARRRAVTLPRPGAPGEAKTSPKD
25 DLDE*

25

15 Seq ID No 98

>Contig12_001 514 amino acids MW=56145 D pI=8.82 numambig=0
PPAVRRYVADRRPEQLPALAPEEREAAARRLSALGAAPPQVRRRGLTRAPLSYQGSRIYF
30 LEQLSPGKPLFNVPGAVRLRGPVDVARLSAAGFEIVRRHDALRTSIANVDGELLQIAQPH
AGFALDVVTSTPEEAAELDRRLRAEAWRPFAIGAPPLLRATLFRLAEDEHVLLVTMHHV
20 SDDWSLGVILRELLALYAGRSLPPRLQVSDFAAWQREMVESGALDGQRAYWRERLRGLS
35 RASISAGGGAEAPSHDPSGAIEEIALSPDKAAALEALARREGATLFMVLLALLDLVIHAR
SGALDIAVGTPIANRRRPELEDVVGLLTNTLVIRVDLARAGAFRDVLARARVQALDAFAN
QDIPFDVVTQDLKQERDHAQHPLFRVWLALQNAPKPALEVRGLRVEPLPLRPELVHFEVA
40 LLLWPADDGSSVVGHFEFRRDRVDEGARKEIAAFAFTHLVDAVIARPDAPVSTLVEGARAEA
25 ARAQAALGEAFARAATARLGQLRRRSAGDRTPRE*

40

Seq ID No 99

>Contig12_009 582 amino acids MW=65555 D pI=8.72 numambig=0
MREPSSTPNWRNFGSNLPAGSDSVPPGEGFPIKKILALNLGKWKDTAGLQIAQALHLFEY
30 GYKRYREGKFVLRATSDLGLGAIFESIDNWESFDQFEEFFKPWTFIRKPLVATRWAEDAE

50

55

5

10

15

20

25

30

35

40

45

50

55

FGRQRLVGINPAHIRRATPADLADFVSGAEPKPIAIADGRTLEEVREGGQLYFLDYRIFK
DIVD TDVQEELGKYPLAPTCMLHQTAAGELLPVAILVHSRPGKGAHPDKIFTSPGPSDD
WLTAKIAVASADAIYQGQVTHLLYAHLIVEPFAVSTYRNLPATHPLHQLLRPHFFNTLAI
NELARRRFLGRGRFFDITSSVATMGSFELLTRAYTGKGIKGYGGKPWRFYESALPRDLA
RDVRLVGYHYRDDALLHWDIAIQEYVGQVLKIAIPTPGSLSSDASLQRWIHELVSQQLGG
MDSLLPPERADQLEKLTSLDDLIAIVTNIIFTATAYHAAVNFGQTDYYTWIPNAQFATYR
SYGDLVNGSEKRQFKPLERLPGRAQSIRQMVLSRSLMGPPLTSESLMTMKCLLQDPAAK
QAFARYRERLAHIEREITERNRAREQPYLYLLPSMVPOSVAI*

10 SEQ ID No 100 (>ORF1)

VSQRTSCYLRGGGVCSMNDAFLALERNERNRPSTVIDLLRQRAEAPARPIYCFLESGDVEAG
ATWVTLREIDERARTVAALLQASGVAPGARALLLYPPGIEYITAFFGCLYAGVRTVPAYPPDL
GRLERTLPRVASIVADARAEEAALTSSAVAGIVASLPASAAAAALQRLRWIATDGSPSPGPIEGP
GAALRPESVAFLOQYTSGSTGEPKGVMLTHGNLLHNSRLIAHGFDLTSPDPVGVWLPPYHDMG
LIGGILQALYRRIRVALMSPLSFLQRPWRWLRAVSALGASVSGGPNFAYDLCVRKSSEERAA
LDLRSWEVAFVGAEPVRADTLDRFARAFVSGFRREAFYPCYGLAEATLIVSGGARAEAPVLA
RLAPEEVELGRAVASAAEGARVFGSGRALDPRAVAIVDPAGNELGPGEIGEIIWVSGPSVAVG
YWGRPEETEATFGATLAGSAPRYLRTGDLGFLRGGELFVVGSRKDLIIILGRNHFPQDIEKT
VESSHRAVRPGCSAAFSVEHEGEERLAVVCEVDPRVAADPREIVAAREAVTAHQVLVAHAVAL
IAPGALPKTSSGKVRRECRRAFLEDALGERHVAFAPELLDDASPPDDAPPETEESGRSLLD
ALRSTLARALRLDAGQIDDALPISRFGDLSLAQVAVELQHAQVQRTGRAIPLTSILRGGSRLRLTR
EITRLDGPSSPRVATPGGAVCADRWGTGRFGSSAISRPMERFTTWAGRSGSVPAFKRVDLRRRA
F

25 SEQ ID No 101 (>ORF2)

VYSSAYVLFVACAGTTRVASAPETAGFPLECVGDDGTVLGPDSFVVGYTQVYVFKKERLNTNP
PIDGFTLKLDGNEVAPGEDGLPVVKRCVRSEEQAQCGGRTEPAEDECTTYEIEAVVPEKAAEV
DEEAAGLGGPPAREAIWVDYYTDGGEFDGARRLVSDTTGASRGNGTTWTPPSEPGRVSLWAV
VHDTRGGASVTRREVQVE

30

5

222

SEQ ID No 102 (>ORF3)

10

VVGTVLSAGTGEPLPDIAVTLVRPDGGREEAKTDQGGKFRFKNLPPGKYRVEVAAAGFEPFAA
EEEIAAGEAIEVRYRISLAAPQDGKAPGIEVTVQGERPPREVTRRTIERREIDRIPGTGGDAL
RSLQSLPGVARSGFGLLIVRGSAPODTLTFVDRTVPPIIYHFGGLSSVVPTEMLEKIDFYPGN

5

FSAVYGRAMGGIVDVGLRSPKQDGKYHGVVQLDLIDGRVLLEGVPVFLKDWTFIAAGRRSWVD

15

AWLGPVLKEAGSSVTQAPVYYDYQFVLEGRPSASERVASFYGSDDAFKITLTDKPPPEDEPALT
GDFGLHTAFQRFQLSYENRIGSRDRLLWSMALGRDIADFEISPLAFNVVSTSLDLRLELSHRF
ARYLTMMVGTDLSSGGVATVNIRAPSQQPAGHPSNQPFSTYPPQDRSFDGAYSRAAYAELEVV
PSPRARIVPGVRVDYALDTQTLDVSPRVNARYDIRSGFPRTTAKGGVGLYYQAPQFAESIEPF

20

10

GNAELKSNRAVHYGLGVEQEITPQIEVTLDDGYFKQLDRLVVFSPKDDYADGTGYAVGGELLL
KYKPDERFFGWAAYTLRSVRKDGPDDEEHLTQFDQTHVLTVLGSLRLGRGWELARFRLVSGN
LQTPYVCDPEEKGCNPNRVNAIYHASSARYSPIPLGGDYSERMPLFHQLDIRADKTWKFKRWQ
LGPLYLDIQNVYNMAAEGISYNFNYTKREYVTGLPFLPTLGLRGDF

25

15 SEQ ID No 103 (>ORF4)

VIAVDMNPEAVDAVKDKTSAAFVGDAIVHKVLEGIGAQYVETAIVTFGEHFEPVSLCVASLVR
MGVRIIARAATDRQADILRAVGATRVIQLETEMGRRVGADITMPLAQDLLDLASHYRVVPWNA
HGPLVGQTLGASKIRQRYRINVLGVRPHTNKRPGDKPRLEAPTPDYVIRDGDTLLLVGDSDDV
SRFVAEVGG

30

20

SEQ ID No 104 (>ORF5)

SGSSGGGSSAEGSRCQPSGGGPHWLEGETVTFPVTASGLALAGDAFEVGPPLPEGAAYDPIA
REVTFSPLDQAAVYDIEIRVAQTSEVGRVKVGADAFADPSNVPVVDPTRYPEEYGLPVLFL
SPVPEDKEYAPATVIYRGHTYAAEAEELRGESSLSYPKRSYTLKFPKDDKFNEPDEAGGFTDRR
KVVLIITTFDDNSYVRQRLAYDLWNRDLPEHIQIKTYSVAVLYLDGEYAGLYTVADHVDGYLMED
HGYPQDGNLYKAVSHDANFALTDRSGDPKDTLHDGFEKKEGAPAEGEPEAFSDLEDLVSFVAE
SDDATFAAEIGSRIDLRDYEDWWIFVTFIVANDSAGKNSYHYRDPADGVFRYAPWDFNASFG
QSWETEREPASDRVDYRDVNNLLFERLLEEPSIGDPLRARYDQVLRGALAEAEIHAIVDGYVER
IDASARRDEARWGEAYRSYEGWSWRDDFTTYEEEIAYLK

35

40

25

45

30 AWISERWQHODELY

50

55

5

10

15

20

25

30

35

40

45

50

55

SEQ ID No 105 (Contig 11 >ORF1)

VLDVWSTSDQVACRLHCAGAGPSASLELRYDASAGARRDAERLAERLAALLEDLSRHPERFVA
QGEVGPGERAEIEAWSRGPAMELPSACALHRWFEERAEQHPDVAVRSEGKSLTYGELERRAN
5 RLASCLRRRGVGLDTIVGVCVPRSEDMVVATLAVLKVGAYLPDHEYPERLAFMMRDARAR
LLVTHDAIADELPTGGWTTLLDAAEAAEIAACSDARPAVSPPPDSGAYVIYTSGSTGTPKGSL
15 ISHRAIVNQMWIQRYWALTADDRVLLKAAFQFDVSVWEIFWPLSFGARTVVARAGGHRDPEY
LRLVRDEGATTAYFVSSMLAAFLGGPEQPPASLRKVLVGGEAVPLDLVRRFYAKHDGDLIN
MYGPSEAAIAVTGCVLPSDPRVTWVPLGAPVANAEVFVLDGAMRRPAIGALGDLYIAGAPLAR
20 GYVGQPLTAERFLPDPCARAAGGRMYRTGDVARFLPDGMLEFQGRSDHQIKLRGHRIELGDV
10 EAQIRRVPGVGQAAVVLREDAPGDARLVAYVVLDGDAAGDAPDVRAGLKASLSAYMIPSSVVR
LYALPMCSERLAFTGSSYAGCLL

SEQ ID No 106 (Contig 11 >ORF2)

15 MSDHEMTGFSLSPPQRAIRALDREAGAPGCRTLAVVAVTGPCDEGRLSAAALALAERHEILRT
RLVEGRARPRRWSASRASRGRQQDDWVCSEAEQGERMSRLVARLSEDRGADDGLRVGLVRVG
30 PEERRLVLAAPAWCVDEESIAPLVRELCASTAGAGAPPEQQYADVAEWLNGMLESADAGDGRR
FWAERRSHFGPPLHLAFSRGGAGAGAGSGRARVDLQGGMAQVERWSSSQVPQRIVLLALWAS
LLWRMSGGNEPEVTVAVRFDGRSLDALAGAVGPFARFLPVRIEISASDTLADVARRLALAEAE
20 AAAHQDAAPGVSHRMSWGLLRGGGRAGAVARRRAGPRARRLEHV

SEQ ID No 107 (Contig 11 >ORF3)

MSRIRAQLGVELPLRALFQGPTVAALAAQVDAARRGEARRREFPPIARIIPRDGPLPLSFAQHR
40 LWFVDQLEPGSPAYNIPFVVVRATGRLDVDALRRSLFEIARRHEALRTTFSARDGVPPFVVAPE
25 ARVPFRMSDLEHLAGEALDAAVSALVLEESLAPFDLSRGPLLVRVIRKRHDEHVIALVVHHV
VFDVWSVGVFVGELAALYGGFAQQQPSRLPELPAQYVDFAAAQRAWLSGEVLEGELELYWTTKL
SGALRRARVPVDHEPAGRRTWRGARRSLDAGAELTRQIKAFCEAEISPFMALLAAYKLVLHQ
45 RTGLEDLVVGTDVANRRNVETEMIGFFVNQLVLRDTCGGDPTFGALVRRVRDVALEAFEHQD
LPFDRLVEALRPKGAVGHVPLFDAKFVMRNVHVPPMKLEGLELEALEGEATTTAFDFVLTVAE

AGGSFRFGVEHSSSELYRAATVDNFLSDYRQILATATARPDTVPVSELRGELERAAAAARRELERK
AARGAALDKLTSARRRAVTLPRPGAPGEAKTSPKDDLDE

SEQ ID No 108 (Contig 11 >ORF5)

MSEPIETEDGGSDIAIVGMAGRFPGAPSVDALWENVRRGVESIAARFPESEREPEPPVGASAAPG
APVVCAGLLDDIDRFDASYFGYS PREAQLMDPQQLFLECAVALEDAGCDPARFPGAIGVF
GGCGSNTYLLQLLSHPDLAATVDPHALMLASEKDYLATRVSYKLDLHGPPSVVVQTACSTSLVA
VHMACESLLGGQCDLALAGGVSIGIPQKRGYPYVPGSICSPDGRCRPFDAEAGTVGGSGVGI
VALKRLADALDRNTVHAVIRGS AVNNDGGRKVGFMAPSVDGQAAAISEAQSVAGVDPGSIGY
VEAHGTATAIGDPIEVEALTQAFRRKTPRKAYCALGSIKANIGHLDAAAGVAGLIKAAHVRS
GEIPPCVHFEAPNPKDLAASPFVPREAAPWPRELPRRAGVSSFGIGGTNAHVVLEEPPPL
PPRAPAPERDHLVTLTARTPEALSTACAQLAAHLEATDVPLDDVAFTLQTGRAEHYPYRAVVA
RTRAEAIQGLAREGASALARPDEPRSSRSRARARRPSGWPARTRRRRSGAPSTRARRRRGR
AASISARSSSARARATGARCSAPRWSRPRSSPSSSRSPGSG

SEQ ID No 109 (Contig 11 >ORF6)

VVDHHVVVEYWSFALIVRELGELYSALRAGRPPQLPPSSFFAAGVSCPSPREAAGGAEYWRK
ALDGTTAIDLPRDRARHDAGARRGRAHAI TLPKPLTGALARLARERGTTLFSVLLSALTVLLH
RASGQSDLVVGVP SAGRHDDESARAFGYFVQMLPVRVALRGAASFDALVARVRDAFLDALAHG
DSALRHLLARAQGEAQRDALFDVAFAPQSTPPSLDARSALAIGVGDVRIAQGELELTTLADEQ
AAAEFDLALFAAELDAGIALRFEYDQQLFDPATIERMARHFVVLLES AVEHPGRPLSELRLMS
DAERALLLDWSGAAAAARQAASAPAPACVHALFEAHAARQPDATLEFGHQRTYAQLSTWST
ELALWLRDRGVGPGSVVGVCIERSPRMVAAQLAVLKAGAAAYASLDPANPPARLAEMLADCRAS
LALTSSQASHKLTAAPCPVHLVQDGACAPSTHILVSRPDDLAYVLFTSGSTGTPKGVCVRHA
SLSRLVSFLHLRLDLSPSDRWTQVASSGFDASVYEIWTPLACGAALLLADDDALRSPTALVSW
LVAQRATLSFMPTPLAEACFEQDWTGSALRAMTVGGDKLHPLRRPPFRLFNMYGPTEATVITT
VAEIADLGAEPPLGRPVDSALVYVLDPHMQPVPPGALGELYIGGACLAQGYTRTDLTAERFLP
DPFGQPGARLYRTGDLVRWRPDGQLAFAGRRDEQVKLRGRRVELGEVESVLRRLPGVREGIVV
LHGQGSAAHLIAHVVPDAHPPSERDLREGMARLVPDALVPAHFVLLPALPMSLSGKVDKLLP
APPAAHAADYEPPSGELELELAHINQSVLHLDRVGRHDSFFDLGGHSLLAMQVLGRIESSLGIR

5

10

15

20

25

30

35

40

45

50

55

TTLRTLFEHPTLAQLASHLSSGAASTSAAAATALERGLTRPDGPSSPRVATPEEPFALTEGQR
AMWLECQKSADG
ALYNLGRTVRLGAGVDVAALRRAFEGGLVERHEALRTTFLTRDGHPLQQVHRHVALEWAEPPAM
ALDEREIVARADEVRRRAFDLERGPLLRVHVWRRGEGQPPLLTVVVHHLVVDYWSFALLVREL
5 GELYSALRAGRPPQLPPSSFFAAGVSCPSPREAAGGAEYWRKALDGATTIDLPDRARHDA
SPRRGRAHAITLPKPLTGALARLARERGTTLFSVLLSALTVLLHRASGQNDLVVGVPSAGRND
15 DESTRAFGYFVQMLPVRVALRGAASFDALVARVRDAFLDGLAHGDSALQHLLAEPRGAARRGG
ALFDVAFAFQGALPSLDPRLAALTGAEDVRIAQGELELTTLADEQAAA EFDLALFAAEELDSG
IALRFEYDQQLFDPATIERMARHFVLLLES AVEHPGRPLSELRLMSDAERALLLDDWSGAAAA
20 RQAASAPACVHALFEAHAARQPDATALEFGHQRTYAELSTWSTELALWLRDRGVGPGSVV
GVC IERSPRMVAAQLAVLKAGAAAYASLDPANPPARLAEMLADCRAALVLTSSQASHKLTAAPC
PVHLVQDGACAPSTH IPLVSRPDDLAYVLFTSGSSGTPKGVCVRHASLSRLVSFFQHLLALSP
RDRWTQLASSGFDA SVYEIWTPLACGAALLLADDDALRSPTALVSWLVAQRATLSFMPTPLAE
25 ACFEQDWTGIALRAMTVGCDKLHPLRRPLPFRLFNMYGPTEATVITTTVAEVADLGDEPPLGRP
IDSALVYVLDPHMQVPVPGVLGELYIGGACLAQGYTRDTLTAERFLPD PFGQPGARLYRTGDL
VRWRPDGQLAFAGRRDEQVKLRGRRVELGEVESALRRLPAVREGVVVLHGQGSAAARLIAYVVP
30 GADPPSERDLREGMARLVPDALVPAHFVLLPALPMSLSGKVDKLLPAPPAAHADYEPPSGEL
ERELAHIWQSVLHLDRVGRHDSFFDLGGHSLLAMQVLGRIESSLGIRTTLRTLFEHPTLHQLA
DRLSSGAASTTAAAATVPASEIAPSLGRAPAD
20 EPYPLSYEQERLWVLEQLLPGGTAYNVVQAVRLRNLVDVDALSSALAALVRRHWSLRTVFVAS
PTPQKICEPEAAPAEVVDLRGTPPDEAEAAAARAWASREQATGFDLARGPVFRARLFRLDHDDVC
VLVLSTHHIVTDAWSFQPLVRDLAELYRRARGGGPADMPPELPLQYVDFAVWQRRHLAGKRLAD
KLAHWTATLRGLPVLELQTDPRPPVQTFRGAERVLPDLARLVAQLDELARSRGATRFMVLLA
40 ALGVLLRRSSGQDDLAIGTAVANRPRPELEPLVGFFVNTIVMRDLGGDPTFEELLSRARKVA
25 LEAFEHQDAPFEKVVEAVNPRRDLRSRPLFQVMLVVQNAPTEALELGEVRIEPLDLPVEATRF
DLRFSVEPRGGRDVISLQYNVDLFDAAITDRMLATMQSVLSRATQDPAQVRALSVAPEDRER
ALVAWNDTAVATPDHLRLEEPFFERAVEQPDAVVD AERRLTYGELARRAEAIAAAAASRSGA
45 TANALVAVVMEKGWEQVA AVLGVLRAGAAAYLPDPRLPEERLRHLLHAEVRLVLTQSAVDGT
IAWPAGIERLAVDADERWREQPVARRPPGGSTDDLAYVIYTSGSTGLPKGVMIDHRGAVNTVL
30 DINRRFDVGPEDRVLALSSLSFDLSVYDVFGTLAAGGAVVIPDRTRASDPGHWR ELVERERV

5

226

10

15

VWNSVPALMEMLMDASPGAGDPALSSLRLVMMSGDWIPLKLPDRIRACRAPRVVS LGGATEA
SIWSIAHPIADVDPAWRSIPYGRPLANQHTYVLDEGLEPCPIGVPGEIHHIGGIGVALGYWRDE
ARTRERFLKHPTTGERLYRTGDLGRYFADGTIELLGRTDHQVKIRGFRIELGEIEAALAQHPS
VEQAVVAAKTDPGSEKRLVAYVVGADGDGAALRDFVRKKLPEYMIPAEVVVLPALPLSANGKV
DRAALPDPAAVAPRAAAVAPRTATERLIASVLAEVLQVEAVGVTDNLFELGFTSLLLVRQRL
LAERIAARAPDEGAAAQAVSLTDLFQYPTIEQLAQRLDAATVKAEPADVGAQRAEARRDARRR
RGRG

20

25

30

SEQ ID No 110 (Contig 12 >ORF1)
PPAVRRYVADRRPEQLPALAPEEREAAAARLSALGAAPPQVRRRGLTRAPLSYGQSRIYFLEQ
LSPGKPLFNVPGAVRLRGPVDVARLSAAPGEIVRRHDLRTSIANVDGELLQIAQPHAGFALD
VVTSTPEEAAELDRRLRAEAWRPFAIGAPPLLRATLFRLAEDHVLLVTMHHVVSDDWSLGI
LRELLALYAGRSLPPRLQVSDFAAQWREMGALDQRAYWRERLRGLSRASISAGGGAEA
PSHDPGATEEIALSPDKAAALEALARREGATLFMVLLALLDLVIHARSGALDIAGTPIANR
NRPELEDVVGLLTNTLVIRVDLARAGAFRDVLARARVQALDAFANQDIPFDVVTQDLKQERDH
AQHPLFRVWLALQNAPKPALEVRGLRVEPLPLRPELVHFEVALLWPADDGSVVGHFERRDR
VDEGARKEIAAAFTHLVDIAVIRPDAPVSTLVEGARAEAAQAALGEAFARAATARLGQLRR
RSAGDRTPRE

35

40

45

SEQ ID No 111 (Contig 12 >ORF2)
MSIHIEEEGRADAAKAPPFQYLQALHSALAHENDPVKRKQIEAGMVFKWLREEPLPFLSQLRR
EKPIFSIPAITLVTRYNDVVEVLNANDVFSVDNIAPKLVENVGQNILAMNDSPKYEHEKSLLR
LAFPRADLPRYRQIVVDEANRLAKVGVDKPFDLTG DYALRVPAGAMARYLG VGEI PTEKVVA
WTHALFNEIFLNPTNDPTAVAAARAARQEALPMIDAIVAARKKQLAKSPPEQPSVLDRYLM
QSVPETYESDEGIRDVILGLLMGCVDLSGGAIVNALVELMKRPRVLRDALNVVNVEDDAAITG
YVLEALRFRPPSTGVTSLCVRDVTGGRGTRHEEKVPAGALVMACSASAMHDHEHIDAPDQFRP
GRLPSRNYLFWESGIHTCHGKYVAILHISLAIKQLLRAGVPSAIDPMPRVHGYPAFPRVRLAA
AEG

50

55

5

227

SEQ ID No 112 (Contig 12 >ORF3)

10

MREPSSTPNWRNFGSNLPAGSDSVPPGEGFPPIKKILALNLGKWKDTAGLQIAQALHLFEYGYK
RYREGKFVLRATSDLGGLGAIFESIDNWESFDQFEEFFK?WTFIRKPLVATRWAEAEFGQRQL
VGINPAHIRRATPADLADFVSGAEPKPIAIADGRTLEEVREGGQLYFLDYRIFKDIVDQVQE
5 ELGKYPLAPTCLMLHQTAAAGELLPVAIRLVHSRPGKGAHPDKIFTSPSGPSDDWLTAKIAVASAD
15 AIYQGQVTHLLYAHLIVEPFAVSTYRNLPATHPLHQLLRPHFFNTLAINELARRRFLGRGRFF
DITSSVATMGSEFELLTRAYTGKGIKGYGGKPPWRFYESALPRDLSARDVRDLVGYYHRDDALLH
WDAIQEYVGQVLKIAYPPTPGSLSSDASLQRWIHELVSQGGMSLLPPERADQLEKLTSLDD
LIAIVTNIIFTATAYHAAVNFGQTDYTWIPNAQFATYRSYGDVLNGSEKRQFKPLERLPGRA
20 QSIQRMVLSRSLSMGPPLTSESLMTMKCLLQDPAKQAFARYRERLAHIEREITERNRAREQP
10 YLYLLPSMVPQSVAI

25

SEQ ID No 113 (Contig 12 >ORF4)

15

VSSSRSTGRVPRDRASPAGSCAPALVPGPPLSYASVMPPLDLHVALFGASGAGKTVLLAAFYR
AQTPQSFQOEYAYKIQAVNKAQGNQLGRFYRLEEGRFPDGSTRFDEYEFDFFPDLPEPAVR
IHWYDYPGRWWEDEPVDADEREAMRQGLIRLGMSQVGILLADGAKYRAEGTGYIRWLFHFAD
30 ECDRLRRASAATGDEVSFPREWILALSKADLCPPDYSARDFEREVCRDADDQLAKLCSVLRAE
HAFGHRFMLLSVAAPAGAQVDPRTSLGVRTLAPAILVSTVEGAVREAQAARKEKSAGETFFQ
GLRDLVQFVDSLDDFLPKRYQIVSKILRFISIKDFATTRLDRLLKMRDAIRKGDFTFAVLTA
20 MVAALRDDEGARAYHQNQ

35

SEQ ID No 114 (Contig 12 >ORF5)

40

MPAPAPLVETSRLLRTRGEHWYEFICVPEIPALPAWLSTLEAMLADADAGAGELRYGLLEI
DDRGQRAPRAYPYVAVRFLDPARRDWTGRQVQHFAAWFPPVPEAVAELPEAVPADWHLRVLD
25 GLAGTYGSGEVFGLEPATIRAWKRSHDESRAARAMAIVKATPPVSLGGGEAAPSRWTRVPTLK
KKPPEPPAAAGLLSVGAVPSGQGRRFGCFAIGAMMLAAFCRLMLACGVRLGGA

45

SEQ ID No 115 (Contig 12 >ORF6)

VRFRSSLGPLLLAALGAALTVSAAFRSAEASVFDSASRPEDADGHVRIPVCIDPTSSAEQRV
30 DGAAGGLIHAPNPSLADVITRVRTALQGSWERWSSVRFTGWESCDSSLPATRMTYGVGRIHPD

50

55

5

10

APNQSDSIGVYNKGGSVQFKPWGADFNRCIKYNWQTARVEYSFDCVEQYAIHEMGAIGFMHE
WHHPLVPSACSQREPLPASDVASGWPPSSRRYIVVNPFGFYDYSIMTYWSGCSDQDGVRFSGSET
LDAVDIQAVATVYPPVGGAPDVCNPGWFAGKRWFCAAQPTVSVGNSCSSGWVECLPHCNPRPF
QGEWWTCTPNPYAVTGQSCSARWELCGD

5

15

SEQ ID No 116 (Contig 12 >ORF7)
VGESQGALVGGNALSTNALNLNALNLNALNLNLSGLSARNLAAIQDPGPGSALARDFLRY
AASCALSSTASFDFSWTDSNGKRHDERYPGLLGVA PAWASGPLDDAGQRIVSSCVAARVNYQ
VPVLLSARSLRDPLKTLSSSQELIDYPDVEGAFWGNLFAAQPYINACYN SATVDNSRAYQRDC
AAGHVTSGGQIVECGLIRIAGSCDRVCQKLNAGAGQYYPSCVDRPGQSTATTKDVITTALP

20

10

SEQ ID No 117 (Contig 12 >ORF8)
VLAHCERGGTLTARAASLLARGAELAAARRAYLDAEGCYGRVEALLGALLPEERRARGLARFR
LGRHTEALADLAAAREAAAAAASEAGAEIELLLDEAMILDWTGEYRAARERVAAAERLAGRVAS
PLLGARLLLGVRSLHRADREDEAAAVLTRAAAQAARLGDEGHETHIIALLLLGFILASLGRV
EEAARDLDAVILSCEERSDLMHLGAALNNRGLARALQGDRAGMIADFERTIALGRELGQPAFE
LVGRYNLAEYLYLMDDLAAARPHARAVQAIAPRCGDRHAPVVVTLIARLRLYQGDEAGARRI
ALRLRAARDDAGCEALKPSEDVLCAMIELATRDDRAAWAALEERSARCSVGQERIEVLEARA
LAALRRGRRADARAQLERALAAASTIPTVMGGRLRRWYAELTRATESDAPDIDLAAAEATFTG
ARAREKVEY

25

15

30

20

35

SEQ ID No 118 (Contig 12 >ORF9)
QAYPDLWAERGRQELWLRQLPPRACAQLAREALGDAADGALIDRLVTQSEGQPPFFLEELIRAT
AEGRGDALPETVVMVQVRLEALAPPARRILRAASVLGEVFWRGAVAHLLGGDEAAPLAEHLS
ALVAGELCVRHREGFRFPGE EYSFRQALLREGAYAQLTKDDRALGHRLAADWLEAAGEADPLV
LAHCERGGTLTARAASLLARGAELAAARRAYLDAEGCYGRVEALLGALLPEERRARGLARFRL
GRHTEALADLAAAREAAAAAASEAGAEIELLLDEAMILDWTGEYRAARERVAAAERLAGRVASP
LLGARLLLGVRSLHRADREDEAAAVLTRAAAQAARLGDEGHETHIIALLLLGFILASLGRVE
EAARDLDAVILSCEERSDLMHLGAALNNRGLARALQGDRAGMIADFERTIALGRELGQPAFEL
VGRYNLAEYLYLMDDLAAARPHARAVQAIAPRCGDRHAPVVVTLIARLRLYQGDEAGARRIA

40

25

45

30

50

55

5

10

LRLRAARDDAGCEALKPSEDVLCAMIELATRDDDRAAWAALEERSARCSVGQERIEVLEARAL
AALRRGRRADARAQLERALAAASTIPTVMGGRLRRWYAELTRATESDAPDIDLAAA EATFTGA
RAREKVEY

15

- 5 14. DNA sequence according to any of claims 1 to 5 wherein the
DNA is selected from the group consisting of

(a) the following DNA sequences:

20

10 Seq ID No 119 (>Contig17)

25

TTACGTTACTCATCCTATCTCGGCACCCTGTGTGGTGATGTGCTCGCC
TCGAGCGCGAGCGGGACGACGTCGGCGCCGCGCTCGGTGAGCGCCGCGC
GAGGGCGCTCGCGAGATCGCTGGCGACGCCGGCCGGGGCCACGACGAGCC
ACGTCCCCGCGACGTCGCCGCGTGACGCGGCGCTCACGGGTCTCCATTG

30

15 ACGCGGTAGCGCCACGCGCCACGGTGCTCTGCTCTCGGCGGCTCCGCCG
CCACGCCGACAGGGCCGGCATGAGGCTCTCGAGGGCCGAGCGCCGCCCGC
TGTCGGCGACGTGGAGCGGTCCGAGAGCGCCGACGTCGCCGCGCTCG
ATGGCTCGCCAGAACCGGTCTCCTCGGCGGACGCTCCCGGCGCCGCGTC
CTCATCGTCCGACGCGTCGCCTGCGTCGAGCCAGAACCGCTCGCGCTGGA

35

20 ACGCGTACGTCGGCAACGTCACGCGGCGCGCCCGAGCGGAGCGAAGAAC
GCACCCAGTCGATGGCGTGCCGCGCGCGTGGAGCTCGCCTGCCGAGAG
GAGGAAGCGCTCGAGGTGCGCTTCGTGCGGCGGAGCGAGGACACCACGG
TCGCATCGCCGTCGATCGACGAGAGCGTCTCGTCGAGCGCGACGGTGAGC

40

25 GCGCGCGTGAGGGCTGACCTCGACGAAGAAGCGGTGGCCGTCGTCGAGCAG
GGCGCGCGTGCGGTGCTCGAAGCGGACGGTGTGGCGCAGGTTTCGGTACC
AGTGGGCGGCGCCGAGGGCCTCGCCATCAAGCCTCTCGCCCGTCACCGCG
GASTAGAGCGGCACGGTCGCCGGGCGCGGCGGATGCCGTCGAGCGCCTC

45

CAGCATCGTCCGCTCGATGGCCTCCACGTGGGCGGAGTGGGAGGCGTACT
CGACGCGGACCTTGCGGGCGAACAGCTGCGCCCGCTCAGCTCTGCGACG
30 AGTCGTCGATAGCGCCGGGTCTCCGAGACGAGGGCCGCGTGAGGGCT

50

55

5

230

10

15

20

25

30

35

40

45

50

55

GTTGATCGCCGCTATCGCCAGGCGTTCGCCCAAGGGCGCAAGGCGCGCCT
CGAGCTCGGCGGTGGTGAGCTCGACGGCGGACATGGCGCCGCGTCCCGCG
AGCTTCGTAATGGCGCGCGAGCGGAGCGCGACGACCCTGGCGGCGTCTTC
TAGCGAGAGCGCGCCCGCGACGTACGCGGCCGCGATCTCGCCCTGGCTGT
5 GGGCGACGACCGCGTTCGGGCGTGACTCCGGCGGCGCGCCAGGTGGCGGCG
15 AGGGCGATCATGACGGCGAACAGCACGGGCTGCACCACGTGACGCGCTC
GAGCATGGGCGCGGCTGCGCTTCGTGCGCCCGAGCACGGCGAGGAGCG
ACCACTCGACGTGCGGCGCCAGGGCGCGCTCGCACGCTCGATCTCGGCC
CGAAAGGCGGGCGAGGAGCGAGCAGAGCGCGGCCATCGATGGCCACTG
20 CGAGCCCTGGCCGGGAAGACGAAGGCGACCTTGCCCGGCGGGAGCGCCT
CGCCCGCGACCGTTCTGCCCCGCGCGCCCTCGGCGAGCGCCGCGAGC
GCCGAGAGCAGCGCGGCGCGATCGTCTGCCACGACGGCGGCGCGACGCTC
25 GAAATGCGACCGCGTGGTTCGCGAGCGACGCCGCGACGTGACGAGGGCGA
CGTCCTCGTGCTCGGCGAGGTGCGCGTGGAGCTTGCCCGCCTGAGCGCGG
15 AGCGCCGCGTCGCTCTTCGCCGAGAGGAGCACCGGCACCGGCGGCGCGAA
GGGCGCGCGGGCGGGCTCCCCGGCTTGGTTCGTGCGCGGCCGCGCGCG
30 GCGCTTCTTCGAGGACCAGTGCAGCTTGGTGCCGAGATCCCGAACGAC
GACACCGCCGCGCGCCGAGGAGACCCGCTGGCTTCCACGGTACCTCCTC
GGTCAAGAGGCGGATCGCGCCGCGACGACCAATCGATGTGCTGCGACGGGC
20 TCGCGGCGTGAGCGTCTTCGGGAGGACGCCGCTCTGCAGCGCGAGCACC
ATCTTGATGACGCCCGGATCCCCGCGGCGGCTGCGTGTGCCGAGGTT
CGACTTTAGGCTCCCGAGCCACAGCGGGCGCTCCTTCGCGTGCGCCGCGC
CGTACGTGCGAAGAGCGCGCGCGCTCGATGGGATCGCCGAGCGTCGTG
40 CCGGTTCCGTGCGCCTCGACGGCGTCGACGTCCGCGGGGCGAGCCCCGC
25 GCTCGCGAGCGCGTCCCGGATCACGCGCTCTTGCGGGGGCCGTTTCGGCG
CCGTGAGCCCTTGGCTCTTGCCGTCCTGGTTGACGGCCGATCCGCGCACG
ATCGCGAGCACGGGGTGCCCGTCTTCCGGGCGTCCGACAGGCGCTCGAG
45 GAGCACTATCCAGCGCCTTCCGACCAGCCCGCGCGTTCGCGTGCGACG
AGAACGACTTGACACGCCCCGTCCGGCGCGCCCGCGTGTGCGCGCTGAAC
30 TCGCCGAAGATCCCGGGGGTCCGCATCACGGTCACGCCGCGGCGAGCGC

5

231

10

GAGCGAGCACTCGCCTCGACGGATGGCGTGGCAGGCGAGGTGGAGCGCGA
CGAGCGACGAGCTGCACGCCGTGTCGACGCT

Seq ID No 120 (>Contig18)

15

5 TTTTAGGANCCCCGACGTGCACGATCGGCTCGCCAACTCGTGGCGCGCC
GGGACTATTTTACCAGCTCGCGTTGCGCGCCGCGGGACCTACGTGCGG
GGCCTCGTCCGCGCCCCGCACGACGGCGCGCGCCCCCGCGTTCGCGCC
GCGTGGGGCGGCGCTCGTCACGGGCGGGACCGGGCGCTCGGGGCGCACG
20 TTGCCCGTTGGTTCGCGCGGATCGGCGCCGAGCACATCGTGCTCGCGAGC
10 CGCCGCGGAGCGCGGCCCGGCGCGGCCGCGCTCGCCGAGGAGCTTTC
GGTGCTCGGCGCGCGCTGACGCTGCTTGGCTGCGACGTCCCGATCGTG
AGGCGGTGCGGGGCTCGTGCACAACGTCAAGGCCGCGGAGCGACGGTG
25 CGCGCCGTGTTCCACGCGGGCGGTGCGATGCACGAGGCGCCGGTCGCCGC
CATGCGTGTGAGGAGCTCGCCGACGCGATCGCCGTGAAGGCCCGCGGCG
15 CGCAGCACCTCCAAGACGTCTTCGCGCAGCGCCCGCTCAACGCGTTTGTC
CTCTTCTCGTCAGAAACGGGTGTGTGGGCGGTGGCCGGAAGGCGCGTA
30 CGCCGCGGCAACGCGTTCTCTGACGCGCTCGCCGAGGCGCGTCGCGCGG
ACGGCCTCGCGGCGACCTCGATCGCGTGGGGCGCGTGGGCGGGCGGCGGA
ATGCTCGCGACCGACGCCGAGCGGCGCTTGAAGCATCGCGGCGTCGCGCC
35 20 GATGGATCCGGAGCTCGCCGTGCGGGCCCTCGCGACGCGCTCGATCACG
CCGAGACGTGCCTCGCCGTGCTGACGTGACTGGGCGCGCTTCGCCCCG
TCGTTGCGCTCGGCGCGTCTCTCGCCGCTCCTCGACGAGCTCGCGGAGGC
GCGATCGGCGCTCGACGCGCTGCGCGAGCCACCGACGACGCGCGCACGG
40 CCGCCGGTCCCAGCCCGCAAGCACGCTGAGGACCACGCTCGCGGCGCTC
25 CCGGAGGGCGAGCGCCACCGCCACCTCCTCGCGCTCGTGCGGACGGAGAC
GGCGGCGGTGCTCGGGCACGCGGACGCGTTCGCGCGTCGAGCCGAACCGCG
GGTTCCTTTGACCTCGGGCTCGACTCGCTCATGTCCGTGAGCTCCGCAGG
45 CGCGTCCAGCGCGGACCGGCATCAAGCTCCCGGCGACGCTCGCGTTCGA
CCACCCGACGCCGAGCGCGCTCGCGAGCAAGGTGCTCGCCGCGATCGTCC
30 TCCACGACGCGACCCCGCGCGCCTCGCCCGCGCGGAGCTCGAGCGCCTC

50

55

5

232

10

GAGGGGATGCTCTCGGCGATCTACGCGGACGAAGCGCTCCGCGACGACCT
CACGGCGCGCCTCCGCGCCTTCCTGGACAAGCGCGCGGTCCGCACCGAAC
GCCCCGACGACGCCGCGTTTCGCCGAGAAGCTCGGCTCCGCGAGCGCCGAC
GAACTCATTGCCTGATCGATCAGAAGCTCGGAGATCGCATCGATGTCGA
5 CCGTTACTAACGACACGCTCACGGAGTACTTGGCGCGCCTCACTCAAGAG
15 CTCCACAGGAGCGAGACGCGCCTGCGTGCGACGGAAGAGAGGCGACATGA
GCCGATCGCCATCGTCGGCCTCGGGCTCCCCCTCCGGGGCGGGATCCACG
ACCGCGACACGCTCTGGACGTTCTCGAGGAGGGCCGCGACGCCATCGCG
CCGATCCTCGCGAGCCGCTGGAACGCGGACGCGACGTACGACCTCGATCC
20 GGACGCCGTCGGCAAGAGCTACGTGCGCGACGCCGCCATGCTCGATCGCG
10 TCGACCTTTTCGACGCCGATTCTTCGGGATCAGCCCGCGCGAGGCGAAG
TACGTCGACCCGACGACCGCCTCTTGCTCGAGACGTCGTGGCAAGCGCT
25 CGAGGACGCGGGGATTGTGCCGGCGTCGCTGCGAGACTCGAAGACCGGCG
TCTTCGTCGGCACGGGCGCGAGCGACTACGCGTTCTCCAGAGCGATCGC
15 GACGCCTCGGAGGCGTACGCGTTTCATGGGGATGATCTCGTCGTTTCGCGGC
GGGCCGCTTCGCGTTCACGCTCGGGCTCCAAGGCCCGCGCTATCGATCG
30 ACACGGCGTGCTCTTCGTCGCTCGTCGCGCTCCACCTCGCGTGCCAGTCG
CTGCGTCAAGGCGAGTGCGACCTCGCGCTCGTCGCGGTGTGCAGGTCAT
GTCGTCGCCGAGGTGTTTCGTGCTGCTCTCGCGCACGCGCGCTCGCGA
35 GCGACGGGCGATCGAAGACGTTCTCGGCGAACGCCGACGGCTATGGCCGC
20 GGCGAAGGCGTCGTCTCCTGGCCGTCGAGCGCCTCCGCGACGCGCGCGC
GAAAGGGCGCCCGATCCTCGCGGTGATCCGCGGCAGCGCGGTGAACCACG
ACGGCACGTCGAGCGGGATCACGGTCCCGAACGGGCCCGCGCAGCAGAAG
40 GTGCTCCGCGCCGCGCTCGACGACGCGCGGCTTGTCGCCCGCGACGTCGA
25 CGTCGTCGAGTGCCACGGCACGGGGACCTCCATCGGCGATCCCATCGAAG
TGAACGCGCTCGCCGCCGTCTACGGGAGGGGCGCCCAAGGACCGCCCG
CTGTTCTCGGGCGCGCTGAAGACCAACATCGGGCACCTCGAGTTCGCGTC
45 GGGCCTCGCCGGCGTCGCGAAGATGGTCGCTCCATGCGCCACGCGACCC
TCCCCGCGACGCTGCACACGAGCCCGCTCAACCCGCTCGTCGACTGGGAC
30 GCGCTCCCGGTGCGCGTCGTCGACGCCGCGCGCCCGTGGACGCGCGCGA

50

55

5

233

10

15

20

25

30

35

40

CGACGGCGCCCCCGGCGCGCCGGCGTCACGGCGATCGTCGAGGAGGCGC
CCGCCGAGCCCGAGCCACGACGCCCCGACGCCGCGCCGCGCTTCCGGCC
GTGCCCCTTCTCCTCTCGGGCAAGACCGACGAGGCGCTGCGCGCGCAGGC
AGCGCGCCTCCACGCGCACTCGCGGGGCGCCCCGACGCGCGGCTCGTCG
5 ACATCGCCGCGTCGCTCGCGACGACGCGCACGCACTTCGATCGACGCGCG
GCCGTCGTCGCGGCGGATCGCGACGAGCTCCTCGGCGCGCTCGACGCGCT
15 CGCGCGCGGCGAGGCAGGCCCGGGGTCGGTCGTCGCGAGCGCGATCCCCG
CCGGCAGGGTCGTGTTCTGTTCCCCGGCCAAGGCTCGCAGTGGGTCGGG
ATGGCGCGCGCTCCTCGCGTCGTCGGTGGTCTTCCGCGACGAGATCGC
10 GGCTGCGAGCGCGCGCTCGCGCCGACGTCGCCTGGTCGCTCGGCGCCG
TTCTCCGGGGCGACGGCGACGAGGCGACGCTCCTCGGCCGCGTCGACGTC
GTGCAGCCGGTCCTCTTCGCCGTCATGGTCGCCCTCGCCGCGCTCTGGCG
CTCGATCGGCGTCACGCCCCGACGCCGTCGTCGGGCACAGCCAAGGCGAGA
25 TCGCCGCCCGCTACGTCGCCGCGCCCTCTCGCTCGAAGACGCCGCCAAG
15 GTCGTCGCGCTGCGCGCACGAGCGCTCACGAAGATCGCGGGGCGCGGGC
GATGGCCGCCGTCGAGCTCGGCGCACGCGACACCGAGGCGCGCCTCGCGC
CGTTCGGCGACGCCATCGCGATCGCGGCGATCAACAGCCCGCGCGCCACG
30 CTCGTCGCGGGCGACACGGACGCGATCGACGCGCTCGTCCGCGACCTCGA
GGCCGCGCAGATCTTCGCGCGGAAGGTGCGTGTGCGACTACGCGTCGCACT
CGGCGCACGTCGAGGCGATCGAGCGCGAGCTCCTCGCGGATCTCGCGGGG
20 ATCGAACCGCGCGCGGGCGCTGTGCCGCTTTACTCCGCGGTGACGGGCGC
GAAGCTCGACGGGAACCGCCTCGACCCCGCGCATTGGTTCCGGAACCTGC
GCTCGACAAAAAATTTGAGGACGCCACGCGCGCGCTCCACGACGACGGC
40 CGCCGGGTATCCTCATNATCNGGGCGTNCAGAGGAGTCGGTATTNCCCC
25 CCCCCGCTTNCCCG,

or their complementary strands,

45

50

55

5

10

(b) DNA-sequences which hybridise under stringent conditions to regions of DNA-sequences according to (a) encoding proteins or to fragments of said DNA-sequences,

15

(c) DNA-sequences which hybridise to the DNA-sequences according to (a) and (b) because of a degeneration of the genetic code,

20

(d) allele variations and mutants resulting by substitution, insertion or deletion of nucleotides or inversion of nucleotide segments of DNA-sequences according to (a) to (c), wherein the variations and mutants offer isofunctional expression products.

25

15. Peptide encoded by a DNA sequence according to claim 14 selected from the group consisting of

30

Seq ID No 121

>Contig17_001 828 amino acids MW=86259 D pI=5.60 numambig=1
MTVMATPGIFGEFSAQHAGAPDGRCKSFSSHANGAGWSEGAGIVLLERLSDARKNGHPVL
20 AIVRGS AVNQDGKSQGLTAPNGPAQERVIRDALASAGLAPADVDVAEHGTGTTLGDP
35 ARALFATYGAHAHAKERPLWLGLSKSNLGHQTQAAAGIGGVIMVLALQSGVLPRTLHAASP
SQHIDWSSGAIRLLTTEVPWKPGGSPRRAAVSSFGISGTNAHVLEEAPRAAGDDQAGE
PARAPFAPPVPVLLSAKSDAALRAQAGXLHAHLAEHEDVALVDVAASLATTRSHFERRAA
40 VVADDRAALLSALAALAEGRAGAGTVAGEALPPGKVAFVFPQGGSQWPSMARALLASSPA
25 FRAEIEACERALAPHVDWSSLAVLGGDEAHAAPMLERVDVVPVLFVMIALATWRAAG
VTPDAVVGHSSQGEIAAAYVAGALSLEDAARVVALRSRAITKLARGGAMSAVELTTAELEA
45 RLAPLGERLAIAAINSPHAALVSGDPGAIDELVAELSGAQLFARKVRVEYASHSAHVEAI
ERTMLEALDGIAPRPATVPLYSAVTGERLDGEALGAHWYRNLRHRTVRFEHATRALLDDG
HRFFVEVSPHPVLTVALDETLLSIDGATVVSSLRRDEGDLERFLLSAGELHARGHAIDW
30 GAFFAPLGARRVTLPTYAFQRRERFWLDAGDASDDEDAAPGASAEETAFWRAIERGDVAAL

50

55

235

SDALHVADSGRRSALESIMPALSAWRRSRREQSTVGAWRYRVEWRPVSAASRGDVAGTWL
VVAPAGVASDLASALAAALTEGADVVPALALEASDITDTGCRDRMSNVX

Seq ID No 122

>Contig18_002 502 amino acids MW=53019 D pI=6.83 numambig=1
FRXPDVHDLRLANLVARRDYFYQLALRAAGTYVRGLVRAPHD GARPPAFAPRGAALVTGGT
GALGAHVARW FARIGA EHVILASRRGAAAPGAAALAEELSVLGARVTLVACDVPDREAVA
GLVRNVKAGGATVRAVFHAGGAMHEAPVAAMRVEELADAI AVKARGAQHLQDVFQQRPLN
AFVLFSSSETGVWGGGRQGAYAAANAF LDALAEARRADGLAATSIAWGAWAGGGMLATDAE
RRLKHRGVAPMDPELAVAALAHALDHAETCLAVADV DWARFAPS FASARPRPLLDELAEA
RSALDALREPPDDARTAAGPEPASTLRTTLAALPEGERHRHLLALVRTETA AVLGHADAS
RVEPNRGFFDLGLDSLMSVELRRRVQRATGIKLPATLAFDHPTPSALASKVLAAIVLHDA
TPRASPAEALERLEGMLSAIYADEALRDDLTARLRAFLDKRAVRTERPDDAFAEKLGSA
SADELIRLIDQKLGDRI DVDRY*

Seq ID No 123

>Contig18_010 840 amino acids MW=88062 D pI=5.74 numambig=6
MSTVTNDTLTEYLRLRLTQELHRSETRLRATEERRHEPIAIVGLGLPFRGGIHDRDTLWTF
LEEGRDAIAPILASRW NADATYDLDFDAVGKSYVRDAAMLDRVDFDADFFGISPREAKY
VDPQHRLLLET SWQALE DAGIVPASLRDSKTGVFVGTGASDYAFLQSDRDASEAYAFMGM
ISSFAAGRLAFTLGLQG PALSIDTACSSSLVALHLACQSLRQGECDLALVAGVQVMSSPE
VFVLLSRTRALASDGRSKTFSANADGYGRGEGVVVLAVERLRDARAKGRPILAVIRGSAV
NHDGTSSGITVPNGPAQQKVLRAALDDARLVPADVDVVECHGTGTSIGDPIEVNALA AVY
GEGRPKDRPLFLGALKTNIGHLEFASGLAGVAKMVASMRHATLPATLHTSPLNPLVDWDA
LPVRVVD AARPWTRRDDGAPRRAGVTAIVEEAPAEPEPTTPDAAPALPAVPVLLSGKTDE
ALRAQAARLHAHLAGRPDARLVDIAASLATTRTHFD RRAAVVAADRDELLGALDARGE
AGPGSVVASAIPAGR VVFVFPQGSQWVG MARALLASSVVRDEIAACERALAPHVAWSL
GAVLRGDGDEATLLGRVDVVPVLFAVMVALAALWRSIGVTPDAVVGH SQGEIAAAYVAG
ALSLEDAAKVVALRARALTKIAGRGAMAAVELGARDTEARLAPFGDAIAIAAINS PRATL
VAGDTDAIDALVRDLEAAQIFARKVRVDYASHSAHVEAIERELLADLAGIEPRAGAVPLY

SAVTGAKLDGNRLDPAHWFRNLRSTKNFEDATRALHDDGRRVSSXSXAXRGVGIXPPRLX
X

16. Recombinant expression vector which comprises a DNA-
sequence according to any of claims 1 to 10, 12 and 14.

17. Procaryotic or eucaryotic cell which has been transfected
or transformed with a DNA-sequence according to any of claims 1
to 10, 12 and 14 or with a recombinant expression vector ac-
cording to claim 16.

18. Cell according to claim 17, wherein the cell is derived
from myxobacteria.

19. Cell according to claim 17, wherein the cell is derived
from a Sorangium strain.

20. Cell according to claim 17, wherein the cell is derived
from Sorangium cellulosum.

21. Cell according to claim 17, wherein the cell is derived
from a Streptomyces strain.

22. Cell according to claim 17, wherein the cell is derived
from Escherichia coli.

23. Process for an enzymatic biosynthesis, mutasynthesis or
partial synthesis of polyketide or heteropolyketide compounds,
wherein a cell according to any of claims 17 to 22 is culti-

5

237

10

vated in a suitable culture medium and the polyketide or heteropolyketide compound is isolated from the medium.

15

24. Process according to claim 23, wherein the polyketide or heteropolyketide compound is an epothilone.

20

25

30

35

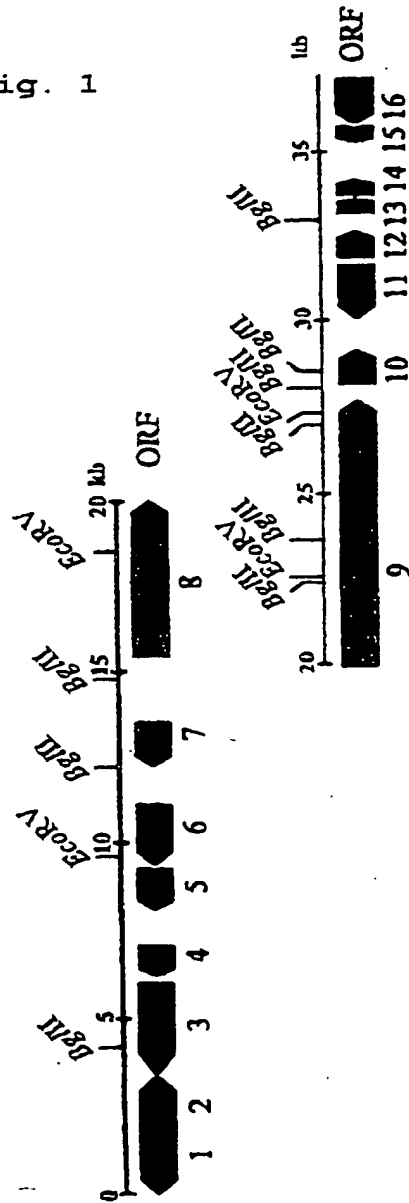
40

45

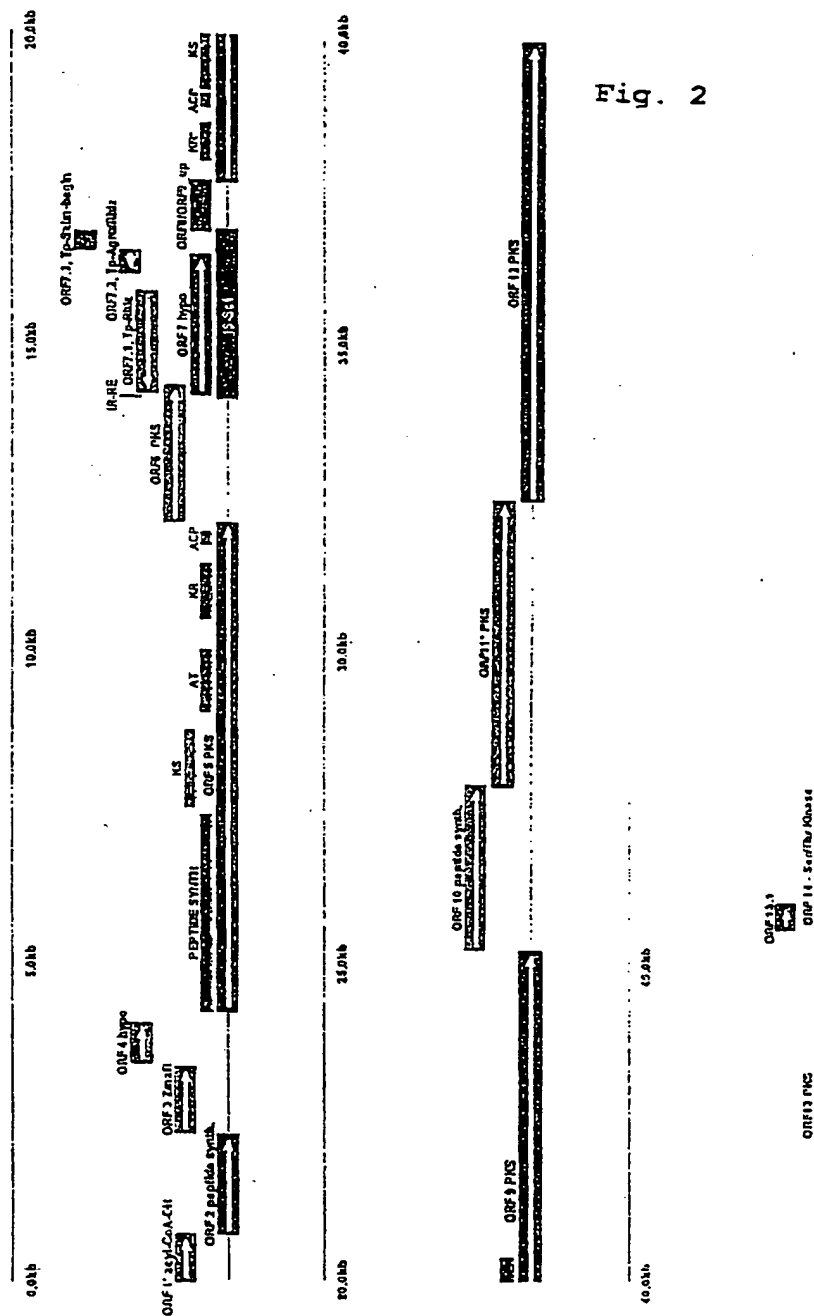
50

55

Fig. 1



Open reading frames found on pEPOcos6 region.
pEPOcos6_annotated
47713 bp



(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
20 April 2000 (20.04.2000)

PCT

(10) International Publication Number
WO 00/22139 A3(51) International Patent Classification⁷: C12N 15/52,
9/00, 15/63, 5/10, C12P 17/06, C07K 14/195, C12P 17/18

(21) International Application Number: PCT/US99/23535

(22) International Filing Date: 11 October 1999 (11.10.1999)

(25) Filing Language: English

(26) Publication Language: English

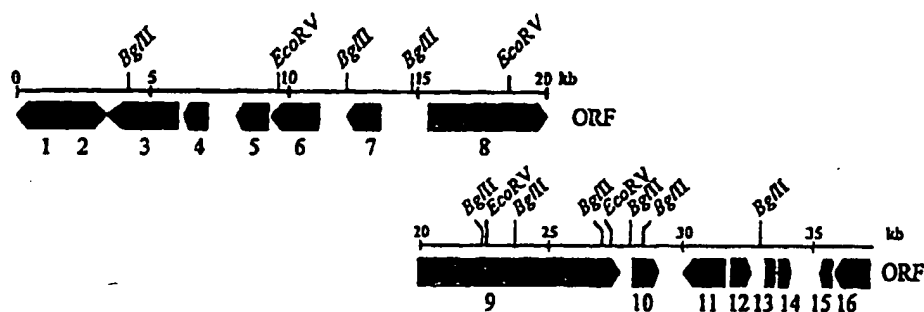
(30) Priority Data:
198 46 493.2 9 October 1998 (09.10.1998) DE(71) Applicants (for all designated States except US):
GESELLSCHAFT FÜR BIOTECHNOLOGISCHE
FORSCHUNG MBH (GBF) [DE/DE]; Mascheroder
Weg 1, D-38124 Braunschweig (DE). BRISTOL-MY-
ERS SQUIBB COMPANY [US/US]; P.O. Box 4000,
Princeton, NJ 08543-4000 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): BEYER, Stefan
[DE/DE]; Mascheroder Weg 1, D-38124 Braunschweig
(DE). BLOECKER, Helmut [DE/DE]; Mascheroder
Weg 1, D-38124 Braunschweig (DE). BRANDT, Petra[DE/DE]; Mascheroder Weg 1, D-38124 Braunschweig
(DE). CINO, Paul, M. [US/US]; 4 Crest Drive, Bound
Brook, NJ 08805 (US). DOUGHERTY, Brian, A.
[US/US]; 10 Rosemary Lane, Killingworth, CT 06419
(US). GOLDBERG, Steven, L. [US/US]; 25 Compton
Court, Basking Ridge, NJ 07920 (US). HOFLE, Gerhard
[DE/DE]; Mascheroder Weg 1, D-38124 Braunschweig
(DE). MUELLER, Rolf-Joachim [DE/DE]; Mascheroder
Weg 1, D-38124 Braunschweig (DE). REICHENBACH,
Hans [DE/DE]; Mascheroder Weg 1, D-38124 Braun-
schweig (DE).(74) Agents: SANTUCCI, Ronald, R. et al.; Kane, Dalsimer,
Sullivan and Levy, LLP 20th floor, 711 Third Avenue, New
York, NY 10017 (US).(81) Designated States (national): AE, AL, AM, AT, AU, AZ,
BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK,
DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,
IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU,
LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT,
RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA,
UG, US, UZ, VN, YU, ZA, ZW.(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent

[Continued on next page]

(54) Title: DNA SEQUENCES FOR ENZYMATIC SYNTHESIS OF POLYKETIDE OR HETEROPOLYKETIDE COMPOUNDS



(57) Abstract: The invention consists of: (1) cloned *Sorangium cellulosum* polyketide synthase (PKS) biosynthetic cluster DNA; and (2) the nucleotide sequence and predicted protein coding sequences of the cloned DNA. The invention can be used for, but not limited to: (a) increasing yields of PKS product in *Sorangium cellulosum* (e.g., by amplification or genetic modification of the epothilone gene cluster or its component parts); (b) increasing yields of polyketide product in a heterologous system by transfer of the epothilone gene cluster or its component parts, which may be followed by amplification or genetic modification of the PKS gene cluster or its component parts; (c) modification of the polyketide product chemical structure in either *Sorangium cellulosum* or a heterologous host (e.g., by genetic modification of the epothilone gene cluster or its component parts; and (d) for the detection of genes and gene products involved in making polyketides or related molecules in other organisms (e.g., by hybridization or complementation assays). DNA sequence and analysis is presented for the following cosmids and plasmids: A2 cosmid; the pEPOcos6 region (overlapping of pEPOcos6 and pEPOcos7); pEPOcos8 cosmid; A5 cosmid; Sau4 (10 kb plasmid).



(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(15) Information about Correction:

Previous Correction:

see PCT Gazette No. 36/2000 of 8 September 2000, Section II

Published:

— *With international search report.*

(88) Date of publication of the international search report:

18 January 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/23535

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/52 C12N9/00 C12N15/63 C12N5/10 C12P17/06
 C07K14/195 C12P17/18

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X A	<p>US 5 716 849 A (SCHUPP THOMAS ET AL) 10 February 1998 (1998-02-10)</p> <p>abstract column 1, line 15 - line 23 column 4, line 8 - line 33 column 4, line 66 - column 5, line 40 example 1</p> <p style="text-align: center;">--- -/--</p>	<p>1,3-5, 16,17, 22,23 18-21</p>



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- *Z* document member of the same patent family

Date of the actual completion of the international search

18 August 2000

Date of mailing of the international search report

23. 10. 00

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
 Fax: (+31-70) 340-3016

Authorized officer

CEDER O.

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/23535

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SCHUPP T ET AL: "SORANGIUM CELLULOSUM (MYXOBACTERIUM) GENE CLUSTER FOR THE BIOSYNTHESIS OF THE MACROLIDE ANTIBIOTIC SORAPHEN A: CLONING, CHARACTERIZATION, AND HOMOLOGY TO POLYKETIDE SYNTHASE GENES FROM ACTINOMYCETES" JOURNAL OF BACTERIOLOGY, vol. 177, no. 13, July 1995 (1995-07), pages 3673-3679, XP000893003 WASHINGTON, DC ISSN: 0021-9193 the whole document ---	1,3-5, 16,17, 22,23
X	WO 93 13663 A (ABBOTT LAB) 22 July 1993 (1993-07-22) abstract ---	1,16,17, 21,23
A	WO 98 22461 A (BIOTECHNOLOG FORSCHUNG GMBH ;GERTH KLAUS (DE); HOEFLE GERHARD (DE)) 28 May 1998 (1998-05-28) the whole document ---	2,24
P,X	BEYER S ET AL: "METABOLIC DIVERSITY IN MYXOBACTERIA: IDENTIFICATION OF THE MYXALAMID AND THE STIGMATELLIN BIOSYNTHETIC GENE CLUSTER OF STIGMATELLA AURANTIACA SG A15 AND A COMBINED POLYKETIDE(POLY)PEPTIDE GENE CLUSTER FROM THE EPOTHILONE PRODUCING STRAIN SORANGIUM CELLULOSUM SO CE90" BIOCHIMICA ET BIOPHYSICA ACTA. GENE STRUCTURE AND EXPRESSION, vol. 1445, no. 2, 14 May 1999 (1999-05-14), pages 185-195, XP000915670 AMSTERDAM ISSN: 0167-4781 the whole document ---	1-5
E	WO 99 66028 A (NOVARTIS ERFIND VERWALT GMBH ;NOVARTIS AG (CH); SCHUPP THOMAS (CH)) 23 December 1999 (1999-12-23) the whole document ---	1-6,8, 16-24
E	WO 00 31247 A (KOSAN BIOSCIENCES INC) 2 June 2000 (2000-06-02) the whole document -----	1-6,8, 16-24

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 99/23535

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-5 and 16-24 partly and 6,8 completely

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: 1-5 and 16-24 partly and 6,8 completely

Nucleic acid and polypeptide sequences involved in polyketide synthesis and related to cosmid A2 and their uses, where the sequences are seq id nos 1,3-34.

2. Claims: 1-5 and 16-24 partly and 7,9 completely

Nucleic acid and polypeptide sequences involved in polyketide synthesis and related to the pEP0cos6 region and their uses, where the sequences are seq id nos 2,35-70.

3. Claims: 1-5 and 16-24 partly and 10,11 completely

Nucleic acid and polypeptide sequences involved in polyketide synthesis and related to the pEP0cos8 region and their uses, where the sequences are seq id nos 71-88.

4. Claims: 1-5 and 16-24 partly and 12,13 completely

Nucleic acid and polypeptide sequences involved in polyketide synthesis and related to the cosmid A5 and their uses, where the sequences are seq id nos 89-118.

5. Claims: 1-5 and 16-24 partly and 14,15 completely

Nucleic acid and polypeptide sequences involved in polyketide synthesis and related to plasmid Sau4 and their uses, where the sequences are seq id nos 119-123.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 99/23535

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5716849 A	10-02-1998	US 5639949 A	17-06-1997
		AU 695364 B	13-08-1998
		AU 2417895 A	04-01-1996
		BG 101107 A	28-11-1997
		BR 9507948 A	18-11-1997
		CA 2192366 A	14-12-1995
		CN 1152941 A	25-06-1997
		CZ 9603601 A	14-01-1998
		EP 0759078 A	26-02-1997
		HU 76534 A	29-09-1997
		WO 9533818 A	14-12-1995
		JP 10501125 T	03-02-1998
		NZ 285166 A	27-04-1998
		PL 317998 A	12-05-1997
		SK 156996 A	06-08-1997
		US 6117670 A	12-09-2000
		US 5670350 A	23-09-1997
		US 5643774 A	01-07-1997
		US 5723759 A	03-03-1998
		US 5679560 A	21-10-1997
		US 5698425 A	16-12-1997
		US 5710031 A	20-01-1998
		US 5686282 A	11-11-1997
		US 5686283 A	11-11-1997
		US 5817502 A	06-10-1998
		ZA 9504686 A	08-12-1995
-----	-----	-----	-----
WO 9313663 A	22-07-1993	CA 2100791 A	18-07-1993
		AU 665526 B	11-01-1996
		AU 1245092 A	03-08-1993
		EP 0626806 A	07-12-1994
-----	-----	-----	-----
WO 9822461 A	28-05-1998	AU 5483798 A	10-06-1998
		BR 9713363 A	25-01-2000
		CN 1237970 A	08-12-1999
		CZ 9901750 A	15-09-1999
		EP 0941227 A	15-09-1999
		NO 992338 A	14-05-1999
		PL 333435 A	06-12-1999
		ZA 9710384 A	18-05-1999
-----	-----	-----	-----
WO 9966028 A	23-12-1999	AU 4611699 A	05-01-2000
		US 6121029 A	19-09-2000
-----	-----	-----	-----
WO 0031247 -A	02-06-2000	AU 1737700 A	13-06-2000
-----	-----	-----	-----